Bayesian Dynamic Modeling of High-Frequency Integer Price Changes*

István Barra[(a)], *Agnieszka Borowska*[(a,b)] and *Siem Jan Koopman*[(a,b,c)]

[(a)] Vrije Universiteit Amsterdam - School of Business and Economics,

[(b)] Tinbergen Institute Amsterdam

[(c)] CREATES, Aarhus University

Abstract

We investigate high-frequency volatility models for analyzing intradaily tick by tick stock price changes using Bayesian estimation procedures. Our key interest is the extraction of intradaily volatility patterns from high-frequency integer price changes. We account for the discrete nature of the data via two different approaches: ordered probit models and discrete distributions. We allow for stochastic volatility by modeling the variance as a stochastic function of time, with intraday periodic patterns. We consider distributions with heavy tails to address occurrences of jumps in tick by tick discrete prices changes. In particular, we introduce a dynamic version of the negative binomial difference model with stochastic volatility. For each model we develop a Markov chain Monte Carlo estimation method that takes advantage of auxiliary mixture representations to facilitate the numerical implementation. This new modeling framework is illustrated by means of tick by tick data for two stocks from the NYSE and for different periods. Different models are compared with each other based on predictive likelihoods. We find evidence in favour of our preferred dynamic negative binomial difference model.

Keywords: Bayesian inference; discrete distributions; high-frequency dynamics; Markov chain Monte Carlo; stochastic volatility.

*Corresponding author: S.J. Koopman, Department of Econometrics, VU Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. E-mail: s.j.koopman@vu.nl IB thanks Dutch National Science Foundation (NWO) for financial support. SJK acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. We are indebted to Lennart Hoogerheide, Rutger Lit, André Lucas, Mike Pitt and Lukasz Romaszko for their help and support in this research project, and to Rudolf Frühwirth for providing the C code for auxiliary mixture sampling. We further like to thank the Editor, Associate Editor and two Referees for their constructive comments. This version: March 21, 2018

1 Introduction

High-frequency price changes observed at stock, futures and commodity markets can typically not be regarded as continuous variables. In most electronic markets, the smallest possible price difference is set by the regulator or the trading platform. Here we develop and investigate dynamic models for high-frequency integer price changes that take the discreteness of prices into account. We explore the dynamic properties of integer time series observations. In particular, we are interested in the stochastic volatility dynamics of price changes within intradaily time intervals. This information can be used for the timely identification of changes in volatility and to obtain more accurate estimates of integrated volatility.

In the current literature on high-frequency returns, price discreteness is typically neglected. However, the discreteness can have an impact on the distribution of price changes and on its volatility; see, for example, Security and Exchange Commission Report (2012), Chakravarty et al. (2004) and Ronen and Weaver (2001). Those assets that have prices with a spread of almost always equal to one tick are defined as large tick assets; see, Eisler et al. (2012). These large tick assets are especially affected by the discreteness through the effect of different quoting strategies on these assets; see the discussions in Chordia and Subrahmanyam (1995) and Cordella and Foucault (1999). Also the effect of liquidity on large tick assets can be substantial as it is documented by O'Hara et al. (2014) and Ye and Yao (2014). Many large tick assets exist on most US exchange markets as the tick size is set to only one penny for stocks with a price greater than 1$ by the Security and Exchange Commission in Rule 612 of the Regulation National Market System. Hence almost all low price stocks are large tick assets. Moreover, many futures contracts are not decimalized for example, five-years U.S Treasury Note futures and EUR/USD futures fall into this category; see Dayri and Rosenbaum (2013).

The relevance of discreteness and its effect on the analysis of price changes have been the motivation to develop models that account for integer prices. Similar to the case of continuous returns, we are primarily interested in the extraction of volatility from discrete price changes. We consider different dynamic model specifications for the high-frequency integer price changes with a focus on the modeling and extraction of stochastic volatility. We have encountered the studies of Müller and Czado (2009) and Stefanos (2015) who propose ordered probit models with time-varying variance specifications. We adopt their modeling approaches as a reference and also use their treatments of Bayesian estimation. The main novelty of our study is the specification of a new model for tick by tick price changes based on the discrete negative binomial distribution which we shall refer to shortly as the ΔNB distribution. The properties of this distribution are explored in detail in our study. In particular, the heavy tail properties are emphasized. In our analysis, we adopt the ΔNB distribution conditional on a Gaussian latent state vector

process which represent the components of the stochastic volatility process. The volatility process accounts for the periodic pattern in high-frequency volatility due to intradaily seasonal effects such as the opening, lunch and closing hours. Our Bayesian modeling approach provides a flexible and unified framework to fit the observed tick by tick price changes. The ΔNB properties closely mimic the empirical stylized properties of trade by trade price changes. Hence we will argue that the ΔNB model with stochastic volatility is an attractive alternative to models based on the Skellam distribution as suggested earlier; see Koopman et al. (2017). We further decompose the unobserved log volatility into intradaily periodic and transient volatility components. We propose a Bayesian estimation procedure using standard Gibbs sampling methods. Our procedure is based on data augmentation and auxiliary mixtures; it extends the auxiliary mixture sampling procedure proposed by Frühwirth-Schnatter and Wagner (2006) and Frühwirth-Schnatter et al. (2009). The procedures are implemented in a computationally efficient manner.

In our empirical study we consider two stocks from the NYSE, IBM and Coca Cola, in a volatile week in October 2008 and a calmer week in April 2010. We compare the in-sample and out-of-sample fits of four different model specifications: ordered probit model based on the normal and Student's t distributions, the Skellam distribution and the ΔNB model. We compare the models in terms of Bayesian information criterion and predictive likelihoods. We find that the ΔNB model is favored for series with a relatively low tick size and in periods of more volatility.

Our study is related to different strands in the econometric literature. Modeling discrete price changes with static Skellam and ΔNB distributions has been introduced by Alzaid and Omair (2010) and Barndorff-Nielsen et al. (2012). The dynamic specification of the Skellam distribution and its (non-Bayesian) statistical treatment have been explored by Koopman et al. (2017). Furthermore, our study is related to Bayesian treatments of stochastic volatility models for continuous returns; see, for example, Chib et al. (2002), Kim et al. (1998), Omori et al. (2007) and, more recently, Stroud and Johannes (2014). We extend this literature on trade by trade price changes by explicitly accounting for prices discreteness and heavy tails of the tick by tick return distribution. These extensions are explored in other contexts in Engle (2000), Czado and Haug (2010), Dahlhaus and Neddermeyer (2014) and Rydberg and Shephard (2003).

The remainder is organized as follows. In Section 2 we review different dynamic model specifications for high-frequency integer price changes. We give most attention to the introduction of the dynamic ΔNB distribution. Section 3 develops a Bayesian estimation procedure based on Gibbs sampling, mainly for the ΔNB case of which the Skellam is a special case. In Section 4 we present the details of our empirical study including a description of our dataset, the data cleaning procedure, the presentation of our estimation results and a discussion of our overall empirical findings. Section 5 concludes.

2 Dynamic models for discrete price changes

We start this section with a discussion of dynamic volatility modeling for high-frequency data. Next, we review models for integer valued variables based on such a dynamic volatility specification. The first group of these models includes the ordered probit models based on normal and Student's t distributions with stochastic volatility. The second group is captured by our novel dynamic negative binomial difference (ΔNB) model with stochastic volatility, which nests the dynamic Skellam model as a special case. We then present the main features of our newly introduced ΔNB model.

2.1 Dynamic volatility specification

To capture the salient empirical features of high-frequency trade by trade price changes such as intradaily volatility clustering and persistent dynamics one typically specifies the following dynamic models for the log volatility h_t:

$$h_t = \mu_h + x_t, \qquad x_{t+1} = \varphi x_t + \eta_t, \qquad \eta_t \sim \mathcal{N}\left(0, \sigma_\eta^2\right), \tag{1}$$

for $t = 1, \ldots, T$, where t is a transaction counter (and not a time index), and where μ_h is the unconditional mean of the log volatility of the continuous returns, x_t is a zero mean autoregressive process (AR) of order one, as denoted by AR(1), with φ as the persistence parameter for the log volatility process and σ_η^2 as the variance of the Gaussian disturbance term η_t. The mean μ_h represents the daily log volatility and the autoregressive process x_t captures the changes in log volatility due to firm specific or market information experienced during the day. The latent variable x_t is specified as an AR(1) process with zero mean; this restriction is enforced to allow for the identification of μ_h.

However, there is also another stylized fact of intradaily price changes which is the seasonality pattern in the volatility process. In particular, the volatility at the opening minutes of the trading day is high, during the lunch-hour it is lowest, and at the closing minutes it is increasing somewhat. We can account for the intradaily volatility pattern by further decomposing the log volatility h_t into a deterministic daily seasonal pattern s_t and a stochastically time varying signal x_t as

$$h_t = \mu_h + s_t + x_t, \qquad E(s_t) = 0, \tag{2}$$

where s_t is a normalized spline function with its unconditional expectation equal to zero. Such a specification allows us to smoothly interpolate different levels of volatility over the day. We enforce the zero mean constraint via a simple linear restriction.

In our model we specify s_t as an intradaily cubic spline function, constructed from piecewise cubic polynomials. More precisely, we adopt the representation of Poirier (1973)

where the periodic cubic spline s_t is based on K knots and the regression equation

$$s_t = w_t \beta, \tag{3}$$

where w_t is a $1 \times K$ weight vector and $\beta = (\beta_1, \ldots, \beta_K)'$ is a $K \times 1$ vector containing the values of the spline function at the K knots. Further details about the spline and the Poirier representation are presented in Appendix B. In our empirical study we adopt a spline function with $K = 3$ knots at {09:30, 12:30, 16:00}. The spline function also accounts for the overnight effect of high volatility at the opening of trading due to the cumulation of new information during the closure of the market. The difference of β_K (market closure 16:00) and β_1 (market opening 9:30) measures the overnight effect in log volatility h_t. This treatment of overnight effects follows Engle (2000) and Müller and Czado (2009); alternatively we can introduce a daily random effect for the opening minutes of trading. For such alternative treatments of intradaily seasonality and overnight effects, we refer to Weinberg et al. (2007), Bos (2008) and Stroud and Johannes (2014).

2.2 Ordered normal stochastic volatility model

In econometrics, the ordered probit model is typically used for the modeling of ordinal variables. But we can also adopt the ordered probit model in a natural way for the modeling of discrete price changes. In this approach we effectively round a realization from a continuous distribution to its nearest integer. The continuous distribution can be subject to stochastic volatility; this extension is relatively straightforward. Let r_t^* be the continuous return which is rounded to $r_t = k$ when $r_t^* \in [k - 0.5, k + 0.5)$. We observe r_t and we regard r_t^* as a latent variable. By neglecting the discreteness of r_t during the estimation procedure, we clearly would distort the measurement of the scaling or variation of r_t^*. Therefore we need to take account of the rounding of r_t by specifying an ordered probit model with rounding thresholds $[k - 0.5, k + 0.5)$. We assume that the underlying distribution for r_t^* is subject to stochastic volatility. We obtain the following specification

$$r_t = k, \quad \text{with probability } \Phi\left(\frac{k + 0.5}{\exp(h_t/2)}\right) - \Phi\left(\frac{k - 0.5}{\exp(h_t/2)}\right), \quad \text{for } k \in \mathbb{Z}, \tag{4}$$

for $t = 1, \ldots, T$, where h_t is the logarithm of the time varying stochastic variance for the standard normal distribution with cumulative density function $\Phi(\cdot)$ for the latent variable r_t^*. The dynamic model specification for h_t is given by (1).

Similar ordered probit model specifications with stochastic volatility are introduced by Müller and Czado (2009) and Stefanos (2015). In the specification of Müller and Czado (2009), the rounding barriers are not necessarily equally spaced but need to be estimated. This further flexibility may improve the model fit compared to our basic

ordered SV model specification. On the other hand, the more flexible model can only be fitted to data accurately when sufficient observations for each possible discrete outcome are available. If you only have a few price jumps of more than, say, $\pm$ 10 tick sizes, it may become a problem to handle such large outcomes.

The basic model specification (1) and (4) accounts for the discreteness of the prices via the ordered probit specification and for intradaily volatility clustering via the possibly persistent dynamic process of x_t. The model can be modified and extended in several ways. First, we can account for the market microstructure noise observed in tick by tick returns (see for example, Aït-Sahalia et al., 2011 and Griffin and Oomen, 2008) by including an autoregressive moving average (ARMA) process in the specification of the mean of r_t^*. In a similar way, we can facilitate the incorporation of explanatory variables such as market imbalance which can also have predictive power. Second, to include predetermined announcement effects, we can include regression effects in the specification as proposed in Stroud and Johannes (2014). Third, it is possible that the unconditional mean μ_h of the volatility of price changes is time varying. For example, we may expect that for larger price stocks the volatility is higher and therefore the volatility is not properly scaled when the price has changed. The time-varying conditional mean of the volatility can be easily incorporated in the model, by specifying a random walk dynamics for μ_h, which would allow for smooth changes in the mean over time. For our current purposes below we can rely on the specification as given by equation (2).

2.3 Ordered t stochastic volatility model

It is well documented in the financial econometrics literature that asset prices are subject to jumps; see, for example, Aït-Sahalia et al. (2012). However, the ordered normal specification, as we have introduced it above, does not deliver sufficiently heavy tails in its asset price distribution to accommodate the jumps that are typically observed in high-frequency returns. To account for the jumps more appropriately, we can consider a heavy tailed distribution instead of the normal distribution. In this way we can assign probability mass to the infrequently large jumps in asset returns. An obvious choice for a heavy tailed distribution is the Student's t-distribution which would imply the following specification,

$$r_t = k, \quad \text{with probability } \mathcal{T}\left(\frac{k+0.5}{\exp(h_t/2)}, \nu\right) - \mathcal{T}\left(\frac{k-0.5}{\exp(h_t/2)}, \nu\right), \quad \text{for } k \in \mathbb{Z}, \tag{5}$$

which effectively replaces model equation (4), where $\mathcal{T}(\cdot, \nu)$ is the cumulative density function of the Student's t-distribution with ν as the degrees of freedom parameter. The model specification for h_t is provided by equation (1) or (2).

The parameter vector of this model specification is denoted by ψ and includes the

degrees of freedom ν, the unconditional mean of log volatility μ_h, the volatility persistence coefficient φ, the variance of the log volatility disturbance σ_η^2, and the unknown vector β in (3) with values of the spline at its knot positions. In case of the normal ordered probit specification, we can rely on the same parameters but without ν. The estimation procedure for these unknown parameters in the ordered probit models are carried out by standard Bayesian simulation methods for which the details are provided in Appendix C.

2.4 Dynamic ΔNB model

Positive integer variables can alternatively be modeled directly via discrete distributions such as the Poisson or the negative binomial, see Johnson et al. (2005). These well-known distributions only provide support to positive integers. When modeling price differences, we also need to allow for negative integers. For example, in this case, the Skellam distribution can be considered, see Skellam (1946). The specification of these distributions can be extended to stochastic volatility model straightforwardly. However, the analysis and estimation based on such models are more intricate. In this context, Alzaid and Omair (2010) advocates the use of the Skellam distribution based on the difference of two Poisson random variables. Barndorff-Nielsen et al. (2012) introduces the negative binomial difference (ΔNB) distribution which has fatter tails compared to the Skellam distribution. Next we review the ΔNB distribution and its properties. We further introduce a dynamic version of the ΔNB model from which the dynamic Skellam model is a special case.

The ΔNB distribution is implied by the construction of the difference of two negative binomial random variables which we denote by NB^+ and NB^- where the variables have number of failures λ^+ and λ^-, respectively, and failure rates ν^+ and ν^-, respectively. We denote the ΔNB variable as the random variable R and is simply defined as

$$R = NB^+ - NB^-.$$

We then assume that R is distributed as

$$R \sim \Delta\text{NB}(\lambda^+, \nu^+, \lambda^-, \nu^-),$$

where ΔNB is the difference negative binomial distribution with probability mass function given by

$$f_{\Delta NB}(r; \lambda^+, \nu^+, \lambda^-, \nu^-) = m \times \begin{cases} d^+ \times F\left(\nu^+ + r, \nu^-, r+1; \tilde{\lambda}^+\tilde{\lambda}^-\right), & \text{if} \quad r \geq 0, \\ d^- \times F\left(\nu^+, \nu^- - r, -r+1; \tilde{\lambda}^+\tilde{\lambda}^-\right), & \text{if} \quad r < 0, \end{cases}$$

where $m = (\tilde{\nu}^+)^{\nu^+}\,(\tilde{\nu}^-)^{\nu^-}$, $d^{[s]} = (\tilde{\lambda}^{[s]})^r(\nu^{[s]})_r \,/\, r!$,

$$\tilde{\nu}^{[s]} = \frac{\nu^{[s]}}{\lambda^{[s]} + \nu^{[s]}}, \qquad \tilde{\lambda}^{[s]} = \frac{\lambda^{[s]}}{\lambda^{[s]} + \nu^{[s]}},$$

for $[s] = +, -$, and with the hypergeometric function

$$F(a, b, c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!},$$

where $(x)_n$ is the Pochhammer symbol of falling factorial and is defined as

$$(x)_n = x(x-1)(x-2)\cdots(x-n+1) = \frac{\Gamma(x+1)}{\Gamma(x-n+1)}.$$

More details about the ΔNB distribution, its probability mass function and properties are provided by Barndorff-Nielsen et al. (2012). For example, the ΔNB distribution has the following first and second moments

$$\mathrm{E}(R) = \lambda^+ - \lambda^-, \qquad \mathrm{Var}(R) = \lambda^+ \left(1 + \frac{\lambda^+}{\nu^+}\right) + \lambda^- \left(1 + \frac{\lambda^-}{\nu^-}\right).$$

The variables ν^+, ν^-, λ^+ and λ^- are treated typically as unknown coefficients.

An important special case of the ΔNB distribution is its the zero mean and symmetric version, which is obtained when $\lambda = \lambda^+ = \lambda^-$ and $\nu = \nu^+ = \nu^-$. The probability mass function for the corresponding random variable R is given by

$$f_0(r; \lambda, \nu) = \left(\frac{\nu}{\lambda + \nu}\right)^{2\nu} \left(\frac{\lambda}{\lambda + \nu}\right)^{|r|} \frac{\Gamma(\nu + |r|)}{\Gamma(\nu)\Gamma(|r| + 1)} F\left(\nu + |r|, \nu, |r| + 1; \left(\frac{\lambda}{\lambda + \nu}\right)^2\right).$$

In this case we have obtained a zero mean random variable R with its variance given by

$$\mathrm{Var}(R) = 2\lambda \left(1 + \frac{\lambda}{\nu}\right). \tag{6}$$

We denote the distribution for the zero mean random variable R by $\Delta\mathrm{NB}(\lambda, \nu)$. This random variable R can alternatively be considered as being generated from a compound Poisson process, that is

$$R = \sum_{i=1}^{N} M_i,$$

where random variable N is generated by the Poisson distribution with intensity

$$\lambda \times (z_1 + z_2), \qquad z_1,\, z_2 \sim \mathrm{Ga}(\nu, \nu), \tag{7}$$

with $\mathrm{Ga}(\nu,\nu)$ being the Gamma distribution, having its shape and scale both equal to ν, and where indicator variable M_i is generated as

$$M_i = \begin{cases} 1, & \text{with probability} \quad P(M_i = 1) = z_1 \,/\, (z_1 + z_2), \\ -1, & \text{with probability} \quad P(M_i = -1) = z_2 \,/\, (z_1 + z_2). \end{cases}$$

We will use this representation of a zero mean ΔNB variable for the developments below and in our empirical study.

In the empirical analyses of this study, we adopt the zero inflated versions of the ΔNB distributions, because empirically we observe a clear overrepresentation of trade by trade price changes that are equal to zero. In the analysis of Rydberg and Shephard (2003) the zero changes are also treated explicitly since they decompose a discrete price change into activity, direction and size. All zero changes are treated as in-activities. This decomposition model is particularly suited for the analysis of micro market structure. In our empirical modeling framework, we concentrate on the extraction of volatility in time series of discrete price changes.

The number of zero price changes are especially high for the more liquid stocks. This is due to the available volumes on best bid prices which are relatively much higher. Hence the price impact of one trade is much lower as a result. The zero inflated version is accomplished by the specification of the random variable R_0 as

$$r_0 \quad = \quad \begin{cases} r, & \text{with probability} \quad (1-\gamma) f_{\Delta\mathrm{NB}}(r; \lambda^+, \nu^+, \lambda^-, \nu^-), \\ 0, & \text{with probability} \quad \gamma + (1-\gamma) f_{\Delta\mathrm{NB}}(0; \lambda^+, \nu^+, \lambda^-, \nu^-), \end{cases}$$

where $f_{\Delta\mathrm{NB}}(r; \lambda^+, \nu^+, \lambda^-, \nu^-)$ is the probability mass function for r and $0 < \gamma < 1$ is treated as a fixed and unknown coefficient. We denote the zero inflated ΔNB probability mass function with f_0.

The dynamic specifications of the ΔNB distributions can be obtained by letting the variables $\nu^{[s]}$ and/or $\lambda^{[s]}$ be time-varying random variables, for $[s] = +, -$. We opt to have a time-varying $\lambda^{[s]}$ since it is more natural for an intensity than a degrees of freedom parameter to vary over time. The dynamic modeling of ν could also be interesting but we leave this for future research. We restrict our analysis to the zero inflated zero mean ΔNB distribution $f_0(r_t; \lambda_t, \nu)$ and we further assume that the degrees of freedom parameters for positive and negative price changes are the same, that is $\lambda_t = \lambda_t^+ = \lambda_t^-$. Taking the above considerations into account, the dynamic ΔNB model can be specified as above but with

$$\lambda_t = \exp(h_t),$$

where h_t is specified as in equation (1) or (2). We recognize that $\exp(h_t/2)$ represents

the standard deviation of the latent variable r_t^* in our ordered probit model specification and here we consider $\exp(h_t)$. However, the variance of a ΔNB random variable relies on λ^2 and is not symmetric in λ. Hence we do not model the standard deviation of a ΔNB random variable as such. Also, due to the asymmetry we require to specify λ_t as a positive variable (enforced by the exponential function). The main reason for a letting λ_t be time varying is to simplify the derivation of the sampling scheme. In particular, the current specification is convenient for deriving the auxiliary mixture sampling method, see Section 3 for details.

2.5 Dynamic Skellam model

The dynamic ΔNB model embeds the dynamic Skellam model as considered by Koopman et al. (2017). It is obtained as the limiting case of letting ν go to infinity, that is $\nu \to \infty$; for a derivation and further details, see Appendix A.

3 Bayesian estimation procedures

Bayesian estimation procedures for the ordered normal and ordered Student's t stochastic volatility models are discussed by Müller and Czado (2009) and Stefanos (2015); their procedures, with some details, are presented in Appendix C.

Here we develop a Bayesian estimation procedure for observations y_t, with $t = 1, \ldots, T$, coming from the dynamic ΔNB model. We provide the details of the procedure and discuss its computational implementation. Our reference dynamic ΔNB model is given by

$$\begin{aligned} y_t &\sim f_0(y_t; \lambda_t, \nu), \qquad \lambda_t = \exp h_t, \\ h_t &= \mu_h + s_t + x_t, \qquad s_t = w_t \beta, \qquad x_{t+1} = \varphi x_t + \eta_t, \end{aligned}$$

where $\eta_t \sim \mathcal{N}\left(0, \sigma_\eta^2\right)$, for $t = 1, \ldots, T$. The details of the model are discussed in Section 2. The variable parameters ν, μ_h, β, φ and σ_η^2 are static while x_t is a latent variable that is modeled as a stationary autoregressive process. The intradaily seasonal effect s_t is represented by a Poirier spline; see Appendix B.

Our proposed Bayesian estimation procedure aims to estimate all static variables jointly with the time-varying signal $h_1, \ldots, h_T$ for the dynamic ΔNB model. It is based on Gibbs sampling, data augmentation and auxiliary mixture sampling methods which are developed by Frühwirth-Schnatter and Wagner (2006) and Frühwirth-Schnatter et al. (2009). At each time point t, for $t = 1, \ldots, T$, we introduce a set of latent auxiliary variables to facilitate the derivation of conditional distributions. By introducing these auxiliary variables we are able to specify the model as a linear state space model with non-Gaussian observation disturbances. Moreover using an auxiliary mixture sampling

procedure, we can even obtain conditionally an approximating linear Gaussian state space model. In such a setting, we can exploit the highly efficient Kalman filtering and smoothing procedures for the sampling of many full paths for the dynamic latent variables. These ingredients are key for a computational feasible implementation of our estimation process.

3.1 Data augmentation: our latent auxiliary variables

We use the following auxiliary variables for the data augmentation. We define N_t as the sum of NB^+ and NB^-, the gamma mixing variables z_{t1} and z_{t2}. Moreover conditional on z_{t1} , z_{t2} and the intensity λ_t, we can interpret N_t as a Poisson process on $[0, 1]$ with intensity $(z_{t1} + z_{t2})\lambda_t$ based on the result in equation (7). We can introduce the latent arrival time of the N_t-th jump of the Poisson process τ_{t2} and the arrival time between the N_t-th and $N_t + 1$-th jump of the process τ_{t1} for every $t = 1, \ldots, T$. The interarrival time τ_{t1} can be assumed to come from an exponential distribution with intensity $(z_{t1} + z_{t2})\lambda_t$ while the N_tth arrival time can be treated as the gamma distributed variable with density function $\text{Ga}(N_t, (z_{t1} + z_{t2})\lambda_t)$. We have

$$\begin{aligned}
\tau_{t1} &= \frac{\xi_{t1}}{(z_{t1} + z_{t2})\lambda_t}, \quad \xi_{t1} \sim \text{Exp}(1), \\
\tau_{t2} &= \frac{\xi_{t2}}{(z_{t1} + z_{t2})\lambda_t}, \quad \xi_{t2} \sim \text{Ga}(N_t, 1),
\end{aligned}$$

where we can treat ξ_{t1} and ξ_{t2} as auxiliary variables. By taking the logarithm of the equations and substituting the definition of $\log \lambda_t$ from equation (2), we obtain

$$\begin{aligned}
-\log \tau_{t1} &= \log(z_{t1} + z_{t2}) + \mu_h + s_t + x_t + \xi^*_{t1}, \quad \xi^*_{t1} = -\log \xi_{t1}, \\
-\log \tau_{t2} &= \log(z_{t1} + z_{t2}) + \mu_h + s_t + x_t + \xi^*_{t2}, \quad \xi^*_{t2} = -\log \xi_{t2}.
\end{aligned}$$

These equations are linear in the state vector, which would facilitate the use of Kalman filtering. However, the error terms ξ^*_{t1} and ξ^*_{t2} are non-normal. We can adopt solutions as in Frühwirth-Schnatter and Wagner (2006) and Frühwirth-Schnatter et al. (2009) where the exponential and the negative log-gamma distributions are approximated by normal mixture distributions. In particular, we can specify the approximations as

$$f_{\xi^*}(x; N_t) \approx \sum_{i=1}^{C(N_t)} \omega_i(N_t)\varphi\left(x, m_i(N_t), v_i(N_t)\right),$$

where $C(N_t)$ is the number of mixture components at time t, for $t = 1, \ldots, T$, $\omega_i(N_t)$ is the weight, and $\varphi(x, m, v)$ is the normal density for variable x with mean m and variance v. These approximations remain to depend on N_t because the log gamma distribution is not canonical and it has different shapes for different values of N_t.

3.2 Mixture indicators for obtaining conditional linear model

Conditionally on N, z_1, z_2, τ_1, τ_2 and $C = \{c_{tj}, t = 1, \ldots, T, j = 1, \ldots, \min(N_t + 1, 2)\}$ we can write the following state space form

$$\begin{aligned}
\underbrace{\tilde{y}_t}_{\min(N_t+1,2)\times 1} &= \underbrace{\begin{bmatrix} 1 & w_t & 1 \\ 1 & w_t & 1 \end{bmatrix}}_{\min(N_t+1,2)\times(K+2)} \underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_t \end{bmatrix}}_{(K+2)\times 1} + \underbrace{\varepsilon_t}_{\min(N_t+1,2)\times 1}, \quad \varepsilon_t \sim \mathcal{N}(0, \mathrm{H}_t) \\
\alpha_{t+1} &= \underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_{t+1} \end{bmatrix}}_{(K+2)\times 1} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & I_K & 0 \\ 0 & 0 & \varphi \end{bmatrix}}_{(K+2)\times(K+2)} \underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_t \end{bmatrix}}_{(K+2)\times 1} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \eta_t \end{bmatrix}}_{(K+2)\times 1}, \eta_t \sim \mathcal{N}(0, \sigma_\eta^2),
\end{aligned}$$

where

$$\underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_1 \end{bmatrix}}_{(K+2)\times 1} \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} \mu_0 \\ \beta_0 \\ 0 \end{bmatrix}}_{(K+2)\times 1}, \underbrace{\begin{bmatrix} \sigma_\mu^2 & 0 & 0 \\ 0 & \sigma_\beta^2 I_K & 0 \\ 0 & 0 & \sigma_\eta^2/(1-\varphi^2) \end{bmatrix}}_{(K+2)\times(K+2)} \right),$$

$\mathrm{H}_t = \mathrm{diag}(v_{c_{t1}}^2(1), v_{c_{t2}}^2(N_t))$ and

$$\underbrace{\tilde{y}_t}_{\min(N_t+1,2)\times 1} = \begin{pmatrix} -\log \tau_{t1} - m_{c_{t1}}(1) - \log(z_{t1} + z_{t2}) \\ -\log \tau_{t2} - m_{c_{t2}}(N_t) - \log(z_{t1} + z_{t2}) \end{pmatrix}.$$

Using the mixture of normal approximation of ξ_{t1}^* and ξ_{t2}^*, allows us to build an efficient Gibbs sampling procedure in which we can efficiently sample the latent state paths in one block using Kalman filtering and smoothing techniques.

3.3 Sampling of event times N_t

The remaining challenge is sampling of N_t as all the other full conditionals are standard. We notice that conditionally on z_{t1} , z_{t2} and the intensity λ_t, the N_t's are independent over time. Using the short-hand notation $v = (v_1, \ldots, v_t)$ for a vector of variables for all the time periods, we can write

$$p(N|\gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s, x, z_1, z_2, y) = \prod_{t=1}^{T} p(N_t|\gamma, \lambda_t, z_{t1}, z_{t2}, y_t).$$

For a given time index t, we can draw N_t from a discrete distribution with

$$\begin{aligned} p(N_t|\gamma,\lambda_t,z_{t1},z_{t2},y_t) &= \frac{p(N_t,y_t|\gamma,\lambda_t,z_{t1},z_{t2})}{p(y_t|\gamma,\lambda_t,z_{t1},z_{t2})} \\ &= \frac{p(y_t|N_t,\gamma,\lambda_t,z_{t1},z_{t2})p(N_t|\gamma,\lambda_t,z_{t1},z_{t2})}{p(y_t|\gamma,\lambda_t,z_{t1},z_{t2})} \\ &= \left[\gamma \mathbf{1}_{\{y_t=0\}} + (1-\gamma)p\left(y_t|N_t,\lambda_t,z_{t1},z_{t2}\right)\right] \times \frac{p(N_t|\gamma,\lambda_t,z_{t1},z_{t2})}{p(y_t|\gamma,\lambda_t,z_{t1},z_{t2})} \end{aligned} \tag{8}$$

The denominator in equation (8) is a Skellam probability mass function with the intensities $\lambda_t z_{t1}$ and $\lambda_t z_{t2}$. To calculate the probability $p\left(y_t|N_t,\lambda_t,z_{t1},z_{t2}\right)$ in the second term in the brackets in (8) we use equation (7), as y_t conditionally on λ_t, z_{t1} and z_{t2} is distributed as a marked Poisson process, with marks given by

$$M_i = \begin{cases} 1, & \text{with} \quad P(M_i=1) = \frac{z_{t1}}{z_{t1}+z_{t2}} \\ -1, & \text{with} \quad P(M_i=-1) = \frac{z_{t2}}{z_{t1}+z_{t2}}. \end{cases}$$

This implies that we can represent y_t as $\sum_{i=0}^{N_t} M_i$, so that

$$p\left(y_t|N_t,\lambda_t,z_{t1},z_{t2}\right) = \begin{cases} 0 \quad , & \text{if} \quad y_t > N_t \quad \text{or} \quad |y_t| \bmod 2 \neq |N_t| \bmod 2, \\ \binom{N_t}{\frac{N_t+y_t}{2}} \left(\frac{z_{t1}}{z_{t1}+z_{t2}}\right)^{\frac{N_t+y_t}{2}} \left(\frac{z_{t2}}{z_{t1}+z_{t2}}\right)^{\frac{N_t-y_t}{2}}, & \text{otherwise.} \end{cases}$$

Then N_t conditionally on z_{t1} , z_{t2} and λ_t is a realization of a Poisson process on $[0,1]$ with intensity $(z_{t1}+z_{t2})\lambda_t$, hence the $p(N_t|\gamma,\lambda_t,z_{t1},z_{t2})$ is the probability of a Poisson random variable with intensity equal to $\lambda_t(z_{t1}+z_{t2})$. We can draw all N_t's in parallel by drawing u, a vector of uniform random variables with $u_t \sim U[0,1]$, and setting

$$N_t = \min\left\{ n : u_t \leq \sum_{i=0}^{n} p(i|\gamma,\lambda_t,z_{t1},z_{t2},y_t) \right\}.$$

3.4 Markov chain Monte Carlo algorithm

To complete our Bayesian specification, we need to specify the prior distributions on the model parameters, which we set as follows:

$$\mu_h \sim \mathcal{N}(0,10), \qquad \beta_i \sim \mathcal{N}(0,1), \qquad \frac{\varphi+1}{2} \sim \mathcal{B}(20,1.5), \tag{9}$$

$$\sigma_\eta^2 \sim \mathcal{IG}(2.5,0.025), \qquad \gamma \sim \mathcal{B}(1.7,10), \qquad \nu \sim \mathcal{G}_{[2:0.2:128]}(15,1.5), \tag{10}$$

for $i = 1, \ldots, K$, where $\mathcal{N}$ is the normal, $\mathcal{B}$ is the beta, $\mathcal{IG}$ is the inverse gamma, and $\mathcal{G}_{[2,2.2,\ldots,128]}$ is the gamma distribution on a grid from 2 to 128 with a resolution of 0.2.

The steps of the MCMC algorithm are outlined below, with more details provided in Appendix D.

1. Initialize μ_h, φ, σ_η^2, γ, ν, C , τ, N, z_1, z_2, s and x.
2. Generate φ, σ_η^2, μ_h, s and x from $p(\varphi, \sigma_\eta^2, \mu_h, s, x|\gamma, \nu, C, \tau, N, z_1, z_2, s, y)$.
   (a) Draw φ, σ_η^2 from $p(\varphi, \sigma_\eta^2|\gamma, \nu, C, \tau, N, z_1, z_2, s, y)$.
   (b) Draw μ_h, s and x from $p(\mu_h, s, x|\varphi, \sigma_\eta^2, \gamma, \nu, C, \tau, N, z_1, z_2, s, y)$.
3. Generate γ from $p(\gamma|\nu, \mu_h, \varphi, \sigma_\eta^2, x, C, \tau, N, z_1, z_2, s, y)$.
4. Generate $C, \tau, N, z_1, z_2, \nu$ from $p(C, \tau, N, z_1, z_2, \nu|\gamma, \mu_h, \varphi, \sigma_\eta^2, x, s, y)$.
   (a) Draw ν from $p(\nu|\gamma, \mu_h, \varphi, \sigma_\eta^2, x, s, y)$
   (b) Draw z_1, z_2 from $p(z_1, z_2|\nu, \gamma, \mu_h, \varphi, \sigma_\eta^2, x, s, y)$.
   (c) Draw N from $p(N|z_1, z_2, \nu, \gamma, \mu_h, \varphi, \sigma_\eta^2, x, s, y)$.
   (d) Draw τ from $p(\tau|N, z_1, z_2, \nu, \gamma, \mu_h, \varphi, \sigma_\eta^2, x, s, y)$.
   (e) Draw C from $p(C|\tau, N, z_1, z_2, \nu, \gamma, \mu_h, \varphi, \sigma_\eta^2, x, s, y)$.
5. Go to 2.

The estimation of s is based on the spline specification $s_t = w_t\beta$ in equation (3) where $K \times 1$ vector w_t can be treated as an exogenous vector and $K \times 1$ vector β is a contains the unknown spline values which are treated as regression coefficients and need to be estimated.

3.5 Simulation study

To validate our estimation procedure for the dynamic Skellam and ΔNB models we independently perform the following experiment 50 times. We simulate $20,000$ observation from the model to be estimated and carry out the MCMC sampling based on $20,000$ draws after the burn-in of $20,000$. The true parameters are set as $\mu = -1.7$, $\varphi = 0.97$, $\sigma_\eta = 0.02$, $\gamma = 0.1$ and $\nu = 10$, which are close to those estimated from real data in our empirical study of Section 4. A single experiment takes approximately 5 hours on a 2.90GHz CPU.

Table 1 presents the posterior means, standard deviations and the 95% credible intervals, averaged over 50 Monte Carlo replications. It also shows the mean inefficiency factors and their standard deviations across the experiments. Figure 1 illustrates the

Parameter	True	Mean	Std	95% range	Mean IF	Std IF
μ	-1.7000	-1.7180	(0.0457)	[-1.8083,-1.6287]	177.2659	(60.5684)
φ	0.9700	0.9701	(0.0041)	[0.9614,0.9774]	354.3270	(101.5417)
σ_η^2	0.0200	0.0200	(0.0032)	[0.0145,0.0270]	552.3604	(153.5302)
γ	0.1000	0.0901	(0.0182)	[0.0526,0.1242]	312.7835	(116.7197)
β_1	1.0652	1.0652	(0.0948)	[0.8790,1.2511]	20.3728	(3.7064)
β_2	-0.8538	-0.8538	(0.0448)	[-0.9420,-0.7662]	22.2124	(4.5653)
ν	10.0000	9.7185	(2.3329)	[5.8333,14.8933]	135.1976	(51.5453)

Table 1: Posterior means, standard deviations (in parentheses), the 2.5%–97.5% quantile ranges (in brackets) and the mean inefficiency factors (IF) and their standard deviations averaged over MC 50 replications, for $M = 20,000$ posterior draws after a burn-in of $20,000$ for $T = 20,000$ observations generated from the ΔNB models.

estimation results on a subsample of the initial 15 Monte Carlo replications, while Figure 2 depicts the posterior densities of the parameters from a single simulation. These results indicate that in our stylized setting, the algorithm can estimate the parameters accurately as the true values are within the highest posterior density regions based on the estimates. The posterior distributions for the autoregressive coefficient φ and for the state variance σ_η^2 seem to be the most challenging to estimate efficiently, as their inefficiency factors are high. Nevertheless, the accuracy of their estimates is satisfactory, with the true values on average being within the 95% credible intervals.

4 Empirical study

In this section we present and discuss the empirical findings from our analyses concerning tick by tick price changes for three different stocks traded at the NYSE, for two different periods. In particular, we analyse the bid prices that correspond to transactions in order to account for bid-ask bounds. We consider two model classes and two models for each class. The first set consists of the ordered probit models with normal and Student's t stochastic volatility. The second set includes the dynamic Skellam and dynamic ΔNB models. The analyses include in-sample and out-of-sample marginal likelihood comparison of the models. Our aims of the empirical study is twofold. First, the usefulness of the ΔNB model on a challenging dataset is investigated. In particular, we validate our estimation procedure and reveal possible shortcomings in the estimation of the parameters in the ΔNB model. Second, we intend to find out what the differences are when the considered models are based on heavy-tailed distributions (ordered t and ΔNB models) or not (ordered normal and dynamic Skellam models). Also, we compare the different model classes: ordered model versus integer distribution model.

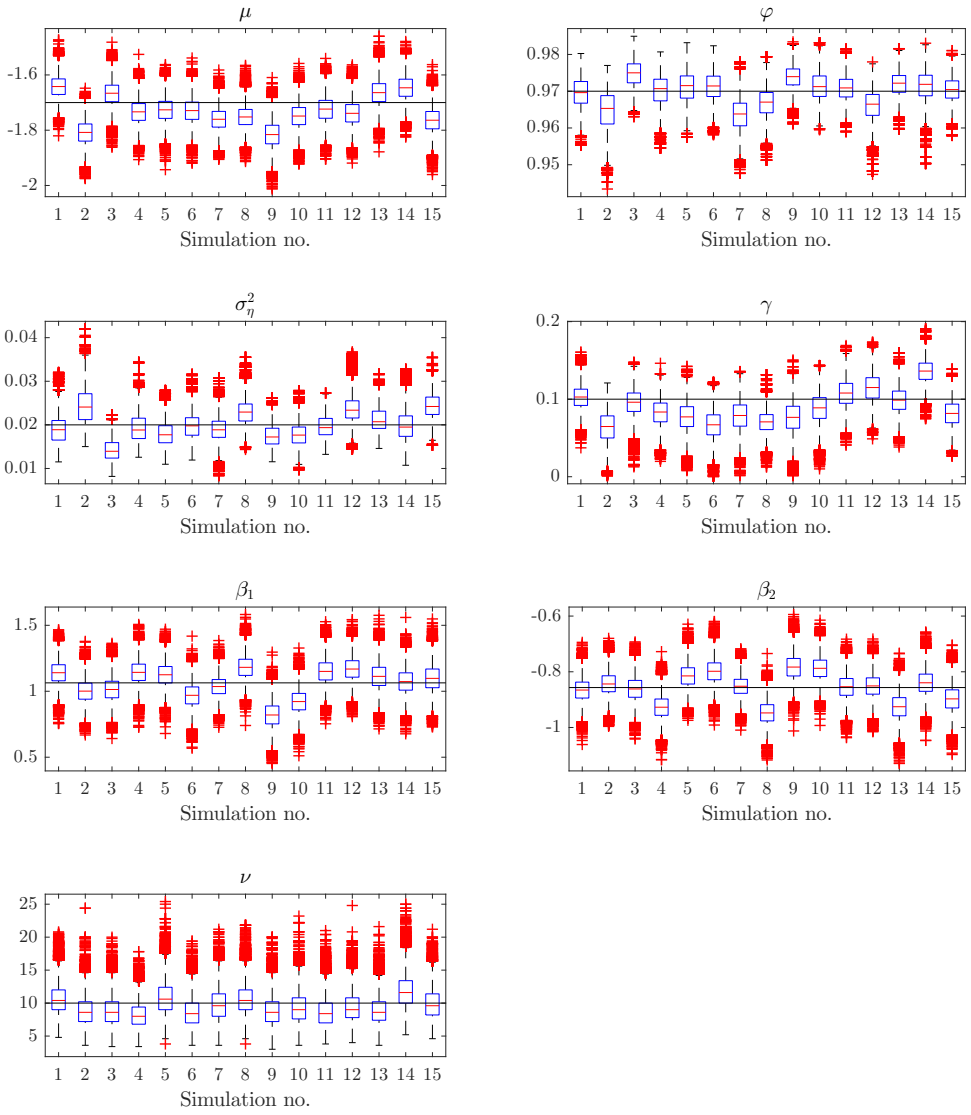


Figure 1: Bar plots of the posterior draws in a subsample of last 15 experiments from our Monte Carlo study.

4.1 Data

We have access to the Thomson Reuters Sirca dataset that contains all trades and quotes with millisecond time stamps for all stocks listed at NYSE. We have collected the data for International Business Machines (IBM) and Coca-Cola (KO). These stocks differ in liquidity and in price magnitude. In our study we concentrate on two weeks of price changes: the first week of October 2008 and the last week of April 2010. These weeks exhibit different market sentiments and volatility characteristics. The month of October 2008 is in the middle of the 2008 financial crises with record high volatilities and some markets experienced their worst weeks in October 2008 since 1929. The month of April 2010 is a much calmer month with low volatilities. To avoid some of the issues related to microstructure noise in high-frequency price changes, including bid-ask bounces, we analyse the bid prices of transactions.

The cleaning process of the data consists of a number of filtering steps that are similar

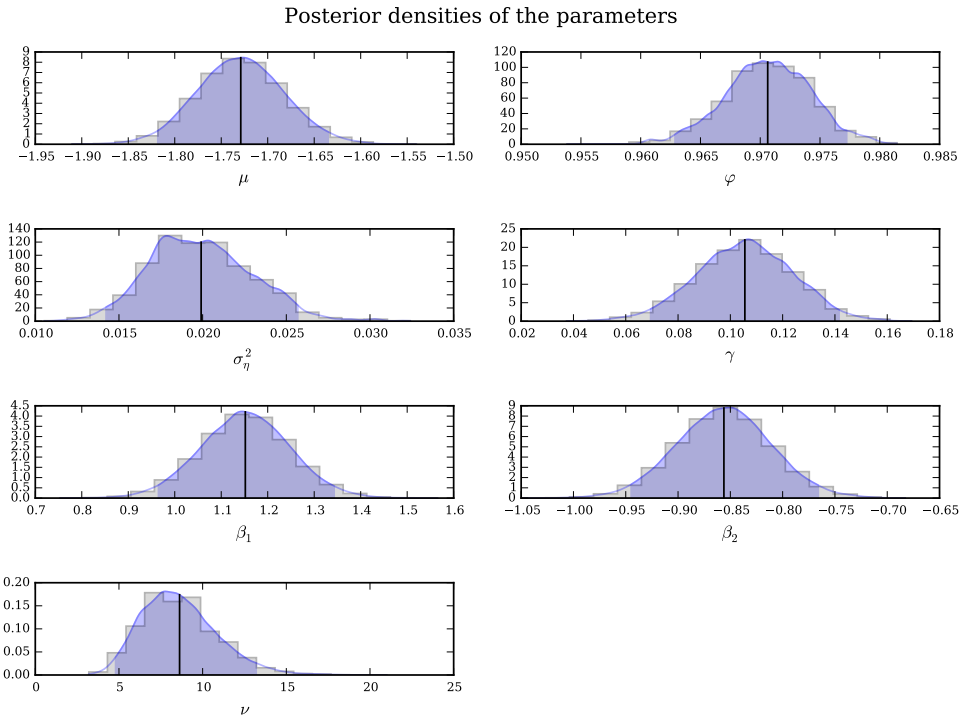


Figure 2: Posterior distributions of the parameters from a dynamic Δ NB model based on 20000 observations and 20000 iterations after a burn-in of 20000. Each picture shows the histogram of the posterior draws the kernel density estimate of the posterior distribution, the HPD region and the posterior mean. The true parameters are $\mu = -1.7$, $\varphi = 0.97$, $\sigma_\eta^2 = 0.02$, $\gamma = 0.1$ and $\nu = 10$.

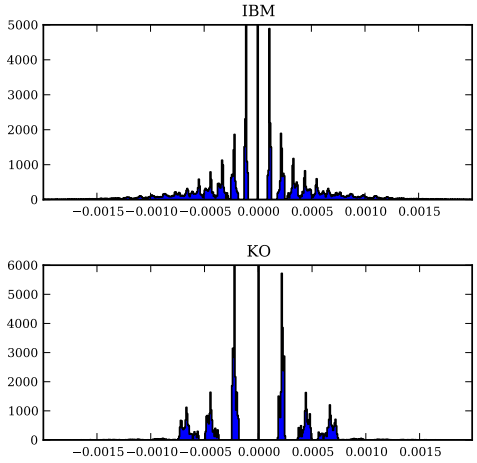
to the procedures described in Boudt et al. (2012), Barndorff-Nielsen et al. (2008) and Brownlees and Gallo (2006). First, we remove all quotes-only entries which is a large portion of the data. By excluding the quotes we lose around $70-90\%$ of the data. In the next step, we delete the trades with missing or zero prices or volumes. We also restrict our analysis to the trading period from 09:30 to 16:00. The fourth step is to aggregate the trades which have the same time stamp. We take the trades with the last sequence number when there are multiple trades at the same millisecond. We regard the last bid price as the bid price that we can observe with a millisecond resolution. Finally, we treat outliers by following the rules as suggested by Barndorff-Nielsen et al. (2008).

Table 2 presents the descriptive statistics for our resulting bid price data from the 3rd to 10th October 2008 and from the 23rd to 30th April 2010, respectively. A more detailed account of the cleaning process can be found in Tables 5 and 6 in Appendix E. We treat the periods from the 3rd to 9th October 2008 and from the 23rd to 29th April 2010 as the in-sample periods. The two out-of-sample periods are 10th October 2008 and 30th April 2010. Figure 3 presents the empirical distributions of the tick by tick log returns as well as the tick returns and the fitted Skellam probability mass function (pmf). For the two stocks considered, IBM and KO, there is a nontrivial number of tick returns higher than 10 in absolute terms. Moreover, we find that the Skellam distribution is too lightly tailed to correctly capture the fat tails of the bid price data.

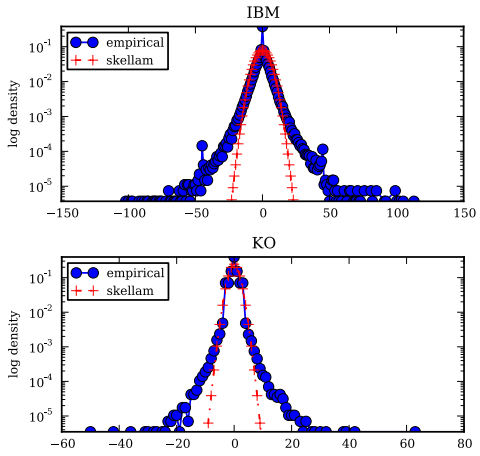
Table 2: Descriptive statistics of the bid prices for IBM and KO from 3rd to 10th October 2008 (top) and from 23rd to 30th April 2010 (bottom). The column "In" displays the statistics on the in-sample period from 3rd to 9th October 2008, while the column "Out" displays the descriptives for the out-of-sample period 10th October. We show the number of observations (Num.obs), average price (Avg. price), mean price change (Mean), standard deviation of price changes (Std), minimum and max integer price changes (Min,Max) as well as the percentage of zero price changes (% 0), the percentage of -1, 1 price changes (%±1) and the percentage of price changes between 2 and 10 in absolute terms (% ± 2–10) in the sample.

October 2008	IBM		KO	
	In	Out	In	Out
Num. obs	68002	20800	70356	25036
Avg. price	96.7955	87.5832	49.2031	41.8750
Mean	-0.0176	-0.0013	-0.0103	0.0046
Std	5.7768	6.3142	1.8334	2.6755
Min	-181	-89	-44	-47
Max	213	169	51	65
% 0	50.1735	48.2981	53.4937	48.2385
% ± 1	8.7233	8.0673	22.0081	17.9022
% ± 2–10	34.4931	34.6346	24.2481	33.0564

April 2010	IBM		KO	
	In	Out	In	Out
Num. obs	43606	8587	34469	6073
Avg. price	130.1758	129.5754	53.6275	53.7317
Mean	0.0014	-0.0181	-0.0029	-0.0061
Std	1.2883	1.3367	0.5971	0.6691
Min	-21	-18	-9	-4
Max	36	10	8	5
% 0	61.6956	60.2888	75.2734	69.3891
% ± 1	23.5472	22.6505	22.2316	26.8730
% ± 2–10	14.6883	17.0374	2.4950	3.7379



(a) Tick by tick log returns.



(b) Tick returns and the fitted Skellam pmf.

Figure 3: Empirical distributions of bid prices for IBM and KO stocks, during the October 2008 period.

4.2 Estimation results

We start our analyses with the dynamic Skellam and ΔNB models for all considered stocks in the periods from 3rd to 9th October 2008 and from 23rd to 29th April 2010. We adopt the same prior specifications as in the simulation study and given by (9)–(10). In the MCMC procedure, we draw $40,000$ samples from the Markov chain and disregard the first $20,000$ draws as burn-in samples. The results of parameter estimation for the 2008 data period are reported in Tables 3a–3b, while for the 2010 period in Tables 4a–4b.

The unconditional mean volatility differ across stocks and time periods. The unconditional mean of the latent state is higher for stocks with higher price and it is higher in the more volatile periods in 2008. These results are consistent with intuition but we should not take strong conclusions from these findings. For example, we cannot compare the estimated means between models as they have somewhat different interpretations in different model specifications. The estimated AR(1) coefficients for the different series range from 0.94 to 0.99, except for the the Skellam and ΔNB models applied to the IBM data in the 2008 period, in which case the posterior means were 0.51 and 0.65, respectively. This finding suggests generally persistent dynamic volatility behavior within a trading day, even after accounting for the intradaily seasonal pattern in volatility. However, by comparing the two different periods, we find that the transient volatility is less persistent in the more volatile crises period. We only included the zero inflation specification for the ΔNB and dynamic Skellam distributions when additional flexibility appears to be needed in the observation density. This flexibility has been required for higher price stocks and during the more volatile periods. The estimates for the zero inflation parameter γ ranges from 0.1 to 0.3. The degrees of freedom parameter ν for the ΔNB distribution is estimated as a higher value during the more quiet 2010 period which suggests that the distribution of the tick by tick price change is closer to a thin tailed distribution during such periods. In addition, we have found that the estimated degrees of freedom parameter is a lower value for stocks with a higher average price.

From a more technical perspective, our study has revealed that the parameters of our ΔNB modeling framework mix relatively slowly. This may indicate that our procedure can be rather inefficient. However, it turns out that the troublesome parameters are in all cases the persistence parameter of the volatility process, φ, and the volatility of volatility, σ_η. It is well established and documented that these coefficients are not easy to estimate efficiently as they have not a direct impact on the observations; see Kim et al. (1998) and Stroud and Johannes (2014). Furthermore, our empirical study has some challenging numerical problems. In the 2008 period we analyse almost $70,000$ observations jointly while the time series in the 2010 period is shorter but still $40,000$ observations. Such long time series will typically lead to slow mixing in a Bayesian MCMC estimation procedures due to highly informative full conditional distributions. Bayesian asymptotic results guarantee

Table 3: Posterior means, standard deviations (Std, in parentheses), the 2.5%–97.5% quantile ranges (95%, in brackets) and the inefficiency factors (IF) for $M = 20,000$ posterior draws after a burn-in of 20,000.

Param.		OrdN	Ord$t$	Sk	ΔNB
μ	Mean	1.1263	1.1052	-0.4523	0.3588
	Std	(0.0329)	(0.0364)	(0.0158)	(0.0332)
	95%	[1.0623,1.1911]	[1.0343,1.1758]	[-0.4836,-0.4216]	[0.2943,0.4246]
	IF	15.4805	25.6141	48.1751	245.6923
φ	Mean	0.9709	0.9763	0.9477	0.9781
	Std	(0.0019)	(0.0025)	(0.0029)	(0.0015)
	95%	[0.9672,0.9747]	[0.9712,0.9810]	[0.9417,0.9537]	[0.9752,0.9807]
	IF	155.0004	762.9785	441.9543	384.1333
σ_η^2	Mean	0.0474	0.0360	0.0329	0.0252
	Std	(0.0036)	(0.0047)	(0.0020)	(0.0017)
	95%	[0.0404,0.0544]	[0.0268,0.0455]	[0.0291,0.0369]	[0.0221,0.0284]
	IF	210.4694	904.4350	775.7113	765.3316
γ	Mean	0.3725	0.3730	0.1896	0.3496
	Std	(0.0030)	(0.0030)	(0.0032)	(0.0032)
	95%	[0.3665,0.3785]	[0.3671,0.3787]	[0.1835,0.1962]	[0.3433,0.3556]
	IF	54.4008	90.2672	239.3389	556.3973
β_1	Mean	0.7383	0.7524	0.3754	0.6454
	Std	(0.0558)	(0.0598)	(0.0260)	(0.0544)
	95%	[0.6300,0.8482]	[0.6359,0.8719]	[0.3245,0.4263]	[0.5411,0.7537]
	IF	4.7332	5.2416	7.0623	30.0869
β_2	Mean	-0.3053	-0.3092	-0.1351	-0.2406
	Std	(0.0557)	(0.0589)	(0.0264)	(0.0534)
	95%	[-0.4148,-0.1978]	[-0.4241,-0.1939]	[-0.1866,-0.0832]	[-0.3464,-0.1372]
	IF	4.2964	4.3154	8.2317	13.7685
ν	Mean		49.9920		17.5968
	Std		(21.6531)		(2.2173)
	95%		[25.0000,111.8500]		[14.2000,23.0000]
	IF		561.1726		839.0705

(b) 2008 KO data.

Param.		OrdN	Ord$t$	Sk	ΔNB
μ	Mean	3.4314	3.3785	0.5416	1.2071
	Std	(0.0338)	(0.0384)	(0.0096)	(0.0094)
	95%	[3.3651,3.4976]	[3.3023,3.4539]	[0.5229,0.5604]	[1.1886,1.2256]
	IF	29.2717	29.2612	385.2575	556.3769
φ	Mean	0.9532	0.9695	0.5101	0.6462
	Std	(0.0024)	(0.0032)	(0.0092)	(0.0083)
	95%	[0.9480,0.9578]	[0.9614,0.9748]	[0.4943,0.5294]	[0.6337,0.6637]
	IF	123.1848	1280.1052	1611.7549	2120.0977
σ_η^2	Mean	0.1259	0.0653	0.5437	0.2280
	Std	(0.0073)	(0.0092)	(0.0082)	(0.0041)
	95%	[0.1118,0.1403]	[0.0505,0.0874]	[0.5278,0.5622]	[0.2191,0.2348]
	IF	216.0429	1520.1055	2003.2192	2653.8330
γ	Mean	0.4500	0.4523	0.3238	0.4190
	Std	(0.0022)	(0.0022)	(0.0025)	(0.0023)
	95%	[0.4457,0.4544]	[0.4478,0.4565]	[0.3194,0.3285]	[0.4144,0.4236]
	IF	17.3129	63.3734	499.1635	72.3874
β_1	Mean	0.5085	0.5148	0.2055	0.2078
	Std	(0.0635)	(0.0691)	(0.0152)	(0.0158)
	95%	[0.3844,0.6330]	[0.3786,0.6505]	[0.1755,0.2349]	[0.1766,0.2389]
	IF	5.6089	5.3234	31.4529	150.0026
β_2	Mean	-0.1528	-0.1628	-0.0629	-0.0785
	Std	(0.0546)	(0.0597)	(0.0136)	(0.0133)
	95%	[-0.2608,-0.0466]	[-0.2804,-0.0440]	[-0.0898,-0.0366]	[-0.1043,-0.0519]
	IF	5.9121	5.1946	48.2788	98.3510
ν	Mean		14.0288		2.2000
	Std		(2.5125)		(0.0000)
	95%		[10.5000,20.6000]		[2.2000,2.2000]
	IF		1218.1696		–

(a) 2008 IBM data.

Param.		OrdN	Ord$t$	Sk	ΔNB
μ	Mean	-1.1537	-1.2438	-2.1331	-2.0890
	Std	(0.0479)	(0.0616)	(0.0564)	(0.0585)
	95%	[-1.2451,-1.0579]	[-1.3648,-1.1230]	[-2.2425,-2.0217]	[-2.2025,-1.9734]
	IF	42.2106	80.5260	20.0063	88.9444
φ	Mean	0.9798	0.9887	0.9926	0.9903
	Std	(0.0031)	(0.0019)	(0.0014)	(0.0017)
	95%	[0.9731,0.9852]	[0.9848,0.9922]	[0.9897,0.9952]	[0.9866,0.9934]
	IF	185.6574	222.6474	284.1253	595.5608
σ_η^2	Mean	0.0187	0.0082	0.0041	0.0065
	Std	(0.0034)	(0.0015)	(0.0007)	(0.0011)
	95%	[0.0129,0.0270]	[0.0057,0.0115]	[0.0028,0.0057]	[0.0046,0.0089]
	IF	236.9756	293.7111	392.8999	914.2363
γ	Mean	0.3608	0.3449	0.0025	0.0224
	Std	(0.0095)	(0.0116)	(0.0019)	(0.0185)
	95%	[0.3417,0.3789]	[0.3220,0.3679]	[0.0003,0.0075]	[0.0003,0.0678]
	IF	136.1222	157.5518	2058.0548	496.8247
β_1	Mean	0.7304	0.7645	0.6191	0.6216
	Std	(0.0707)	(0.0832)	(0.0917)	(0.0833)
	95%	[0.5916,0.8713]	[0.6064,0.9319]	[0.4508,0.8125]	[0.4661,0.7938]
	IF	9.5713	12.5510	50.9314	50.0410
β_2	Mean	0.3251	0.2992	0.4172	0.4261
	Std	(0.0780)	(0.0904)	(0.0973)	(0.0939)
	95%	[0.1716,0.4766]	[0.1230,0.4773]	[0.2228,0.6084]	[0.2404,0.6105]
	IF	4.5545	4.2750	8.2049	9.4661
ν	Mean		17.4954		8.4201
	Std		(3.6575)		(1.8791)
	95%		[12.3000,26.6000]		[5.6000,12.8000]
	IF		389.5012		246.6756

(b) 2010 KO data.

Param.		OrdN	Ord$t$	Sk	ΔNB
μ	Mean	0.5588	0.3051	-0.7590	-0.3193
	Std	(0.0429)	(0.0801)	(0.0298)	(0.0521)
	95%	[0.4750,0.6430]	[0.1500,0.4654]	[-0.8169,-0.6999]	[-0.4184,-0.2130]
	IF	60.3543	14.1064	74.1764	337.6589
φ	Mean	0.9791	0.9959	0.9862	0.9923
	Std	(0.0030)	(0.0009)	(0.0021)	(0.0014)
	95%	[0.9734,0.9842]	[0.9942,0.9975]	[0.9814,0.9902]	[0.9897,0.9950]
	IF	556.1437	175.6690	683.1293	1071.2396
σ_η^2	Mean	0.0234	0.0035	0.0050	0.0042
	Std	(0.0039)	(0.0006)	(0.0008)	(0.0008)
	95%	[0.0168,0.0320]	[0.0024,0.0048]	[0.0035,0.0068]	[0.0027,0.0058]
	IF	650.2627	278.8649	774.4566	1359.5545
γ	Mean	0.4473	0.4317	0.2776	0.3879
	Std	(0.0040)	(0.0042)	(0.0055)	(0.0068)
	95%	[0.4395,0.4552]	[0.4236,0.4397]	[0.2677,0.2882]	[0.3754,0.4021]
	IF	146.9604	52.4582	573.3739	915.2304
β_1	Mean	0.4733	0.6502	0.2909	0.4351
	Std	(0.0711)	(0.1283)	(0.0494)	(0.0814)
	95%	[0.3353,0.6137]	[0.4141,0.9189]	[0.1964,0.3897]	[0.2829,0.6037]
	IF	5.6478	25.0358	12.4851	134.3036
β_2	Mean	0.3954	0.3880	0.2743	0.3286
	Std	(0.0729)	(0.1358)	(0.0512)	(0.0831)
	95%	[0.2512,0.5374]	[0.1179,0.6506]	[0.1733,0.3753]	[0.1618,0.4893]
	IF	4.2390	3.7980	7.1545	9.8248
ν	Mean		7.4336		3.8353
	Std		(0.4564)		(0.4580)
	95%		[6.6000,8.4000]		[3.2000,4.8000]
	IF		219.8048		947.4361

(a) 2010 IBM data.

Table 4: Posterior means, standard deviations (Std, in parentheses), the 2.5%–97.5% quantile ranges (95%, in brackets) and the inefficiency factors (IF) for $M = 20,000$ posterior draws after a burn-in of 20,000.

that long series are more informative about the parameters. Hence we can estimate these parameters accurately. Our Monte Carlo study in the previous section has shown that our algorithm is successful in capturing the true parameters. However, with long series it can be hard to construct efficient proposal distributions. In other words, it can be hard to choose "plausible" parameters in the random walk MH algorithm. We therefore observe low acceptance rates and thus high inefficiency factors. We have also anticipated that parameter estimation for the dynamic Skellam and ΔNB models is less numerically efficient and overall more challenging when compared to the estimation for ordered normal and ordered t models. The estimation for discrete distribution models requires more auxiliary variables and the analysis is based on additional conditional statements.

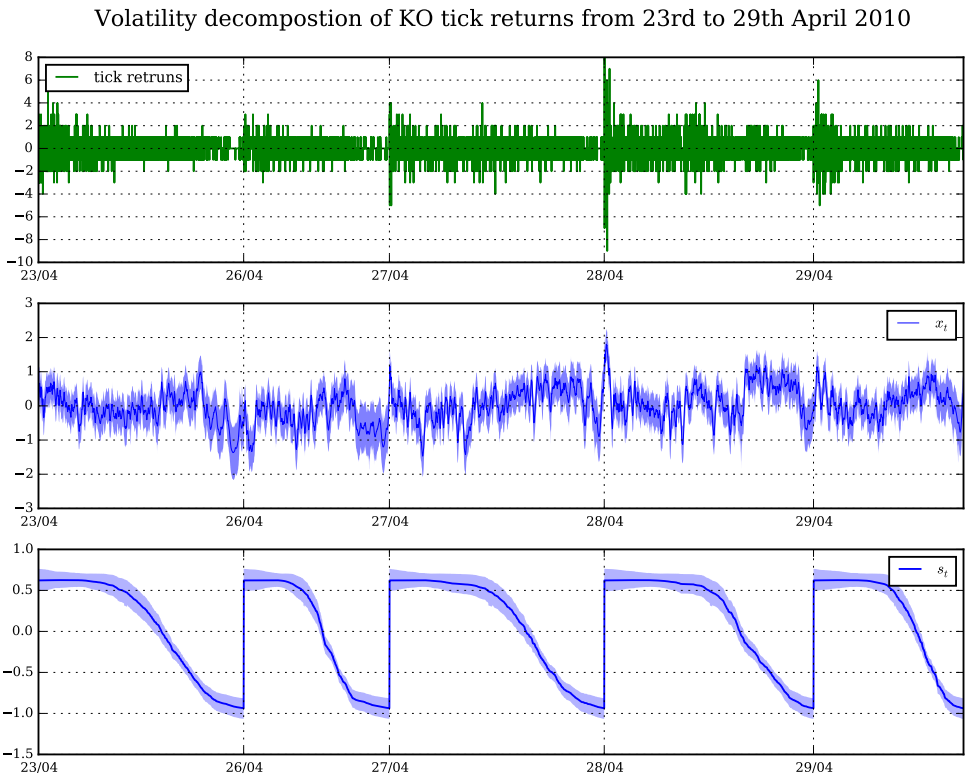


Figure 4: Decomposition of log volatility in the dynamic ΔNB model for KO from 23rd to 29th April 2010.

On the basis of the output of our MCMC estimation procedure, we obtain the estimates for the latent volatility variable h_t but we can also decompose these estimates into the corresponding components of h_t, these are μ_h, s_t and x_t; see equation (2). Figure 4 presents the intraday, tick by tick, Coca Cola bid price changes and its estimated components x_t and s_t for the log volatility h_t in the ΔNB model, from 23rd to 29th April 2010. The intraday seasonality matches with the typical features of tick by tick data and reflects the market mechanism; see also the discussion in Andersen (2000). The volatility is the highest at the beginning of the trading day which is the result of the overnight effect and a different trading mechanism at the *pre-open call auction* during the first half hour of

trading (from 9:00 to 9:30). This burst of information accumulated during the overnight period leads to much higher volatility at the opening of the market. This effect is captured by the estimated initial value of the spline function β_1. The overnight effect receives strong support from the data given that the posterior means are far from zero. We further find that the regular trading takes place continuously throughout the day while it becomes more intense shortly before the closing of the market. The smoothness of the intraday seasonal pattern estimates is enforced through the spline specification. Apart from the pronounced intraday seasonality, we observe many volatility changes during a trading day. Some of these volatility changes may have been sparked by news announcements while others may have occurred as the result of the trading process. Finally, the parameter values underlying the signal extraction of $h_t = \mu_h + s_t + x_t$ are estimated jointly for five consecutive days. Hence it is implied that the overnight effect is the same for each day in our analysis. In Appendix F we compare our estimates of $h_t = \mu_h + s_t + x_t$ with those based on parameter estimates obtained for each day separately. Although some differences are clearly visible, overall the extracted signals of h_t are very similar.

4.3 In-sample comparison

It is widely established in Bayesian studies that the computation of sequential Bayes factors (BF) is infeasible in this framework as it requires sequential parameter estimation. The sequential estimation of the parameters in our model is computationally prohibitive given the very high time dimensions. To provide some comparative assessments of the four models that we have considered in our study, we follow Stroud and Johannes (2014) and calculate Bayesian Information Criteria (BIC) for model $\mathcal{M}$ as

$$BIC_T(\mathcal{M}) = -2\sum_{t=1}^{T} \log p(y_t|\hat{\theta}, \mathcal{M}) + d_i \log T,$$

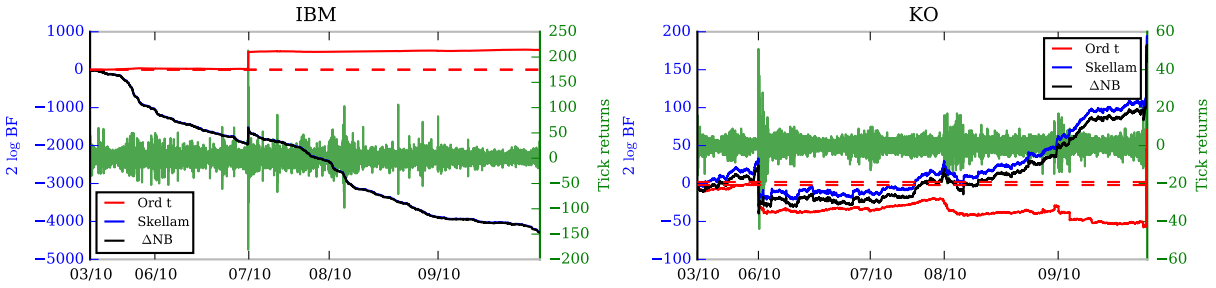
where $p(y_t|\theta, \mathcal{M})$ can be calculated by means of a particle filter and $\hat{\theta}$ is the posterior mean of the parameters. The implementation of the particle filter for all considered models is rather straightforward given the model details provided in Section 2. The BIC gives an asymptotic approximation to the BF by

$$BIC_T(\mathcal{M}_i) - BIC_T(\mathcal{M}_j) \approx -2 \log BF_{i,j}.$$

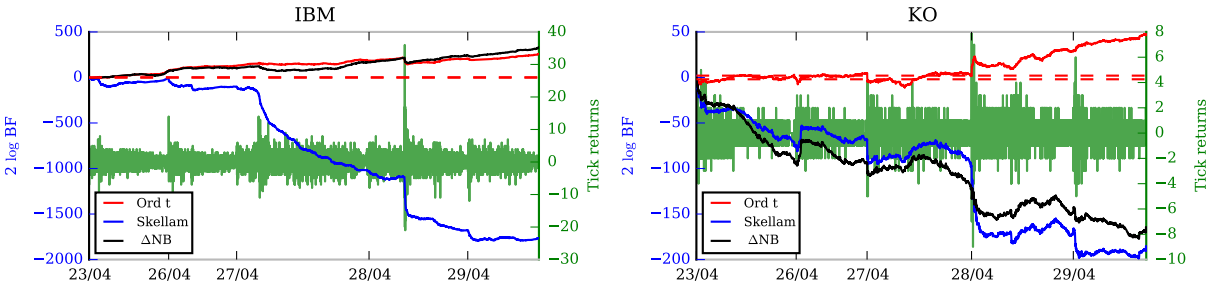
We will use this approximation for our sequential model comparisons.

Figure 5a and Figure 5b present the BICs for the periods from the 23rd to 29th October 2008 and from the 3rd to 9th April 2010, respectively. For the 2008 period, the IBM stock does not appear to favour the integer-based models and its behavior is captured best with the ordered t model. However, the opposite is the case for KO: both

the Skellam and the ΔNB model outperform the ordered models convincingly. In the 2010 period, the IBM stock slightly favours the ΔNB model compared to the ordered t model. In this case the Skellam model does not seem to be able to correctly capture the features of the data. For KO in the same period, the ordered t model provides the best fit to the data, with both the Skellam model and the ΔNB model not performing so well. Furthermore, we may conclude from the BIC results that the ordered t and ΔNB model tends to be favoured when large jumps in volatility have occurred. Such large price changes may lead to a prolonged period of high volatility which suggests the need of the ΔNB model. These findings are consistent with the intuition that for time varying volatility models, the identification of parameters determining the tail behaviour requires extreme or excessive observations in combination with low volatility.



(a) From 3rd to 9th October 2008.



(b) From 23rd to 29th April 2010.

Figure 5: In-sample analysis: sequential BF approximations based on BIC, relative to the ordered normal model, for IBM (left) and KO (right) on two periods.

4.4 Out-of-sample comparisons

The performances of the dynamic Skellam and ΔNB models can also be compared in terms of predictive likelihoods. The one-step-ahead predictive likelihood for model $\mathcal{M}$ is

$$p(y_{t+1}|y_{1:t},\mathcal{M}) =$$
$$\int\int p(y_{t+1}|y_{1:t},x_{t+1},\theta,\mathcal{M})p(x_{t+1},\theta|y_{1:t},\mathcal{M})\mathrm{d}x_{t+1}\mathrm{d}\theta =$$
$$\int\int p(y_{t+1}|y_{1:t},x_{t+1},\theta,\mathcal{M})p(x_{t+1}|\theta,y_{1:t},\mathcal{M})p(\theta|y_{1:t},\mathcal{M})\mathrm{d}x_{t+1}\mathrm{d}\theta.$$

Generally, the h-step-ahead predictive likelihood can be decomposed to the sum of one-step-ahead predictive likelihoods via

$$\begin{aligned} p(y_{t+1:t+h}|y_{1:t},\mathcal{M}) &= \prod_{i=1}^{h} p(y_{t+i}|y_{1:t+i-1},\mathcal{M}) \\ &= \prod_{i=1}^{h} \int\int p(y_{t+i}|y_{1:t+i-1},x_{t+i},\theta,\mathcal{M}) \\ &\quad \times p(x_{t+i}|\theta,y_{1:t+i-1},\mathcal{M})p(\theta|y_{1:t+i-1},\mathcal{M})\mathrm{d}x_{t+i}\mathrm{d}\theta. \end{aligned}$$

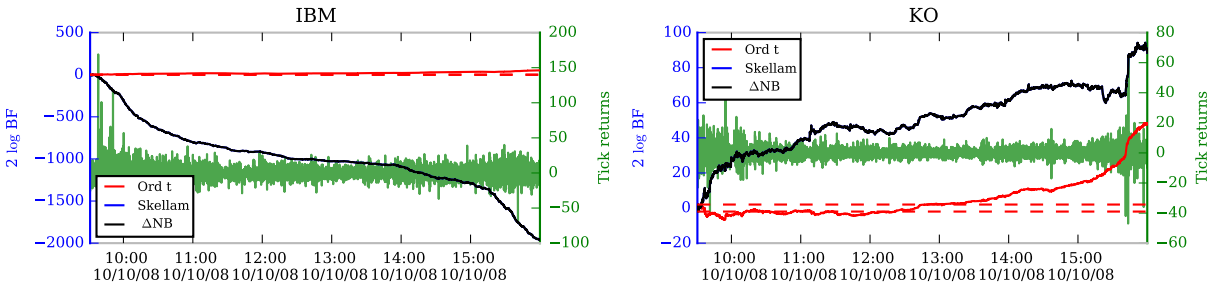
These results suggest that we require the computation of $p(\theta|y_{1:t+i-1},m)$, for $i = 1, 2, \ldots$, that is the posterior of the parameters using sequentially increasing data samples. It requires the MCMC procedure to be repeated as many times as we have number of out-of-sample observations. In our application, for each stock and each model, it implies several thousands of MCMC replications for a predictive analysis of a single out-of-sample day. This exercise is computationally not practical or even infeasible. However, we may be able to rely on the approximation

$$\begin{aligned} p(y_{t+1:t+h}|y_{1:t},\mathcal{M}) \approx \prod_{i=1}^{h} \int\int & p(y_{t+i}|y_{1:t+i-1},x_{t+i},\theta,\mathcal{M}) \\ & \times p(x_{t+i}|\theta,y_{1:t+i-1},\mathcal{M})p(\theta|y_{1:t},\mathcal{M})\mathrm{d}x_{t+i}\mathrm{d}\theta. \end{aligned}$$

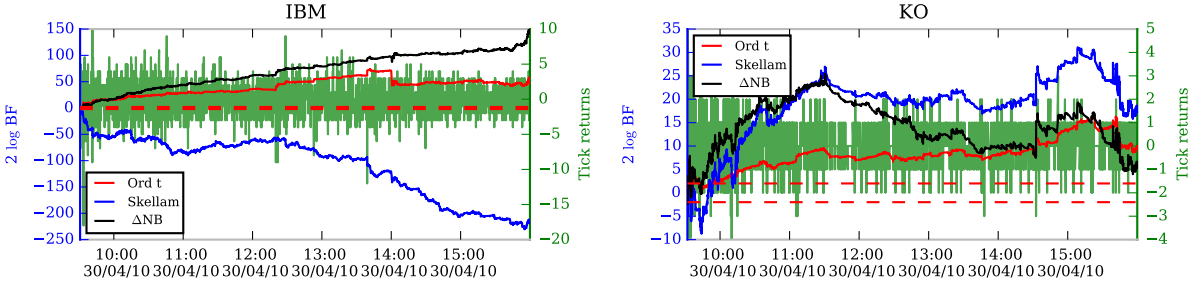
This approximation is based on the notion that, after observing a considerable amount of data, that is for t sufficiently large, the posterior distribution of the static parameters should not change much and hence $p(\theta|y_{1:t+i-1},\mathcal{M}) \approx p(\theta|y_{1:t},\mathcal{M})$.

Based on this approximation, we carry out the following exercise. From our MCMC output we obtain a sample of posterior distributions based on the in-sample observations. For each parameter draw from the posterior distribution we estimate the likelihood using the particle filter for the out-of-sample period.

Figures 6a and Figure 6b present the out-of-sample sequential predictive Bayes factors approximations for the 10th October 2008 and 30th April 2010, respectively. Similarly as in the in-sample 2008 period, also on the 10th October 2008, the ordered t model performs best for the IBM stock, while both integer distribution models perform well for the KO stock. On the 30th April 2010 the ΔNB model performs the best for IBM while the Skellam model is being the worst. It suggests that the IBM stock requires a heavy-tailed distribution, as in the ΔNB and ordered t model. For KO in the same period, both the dynamic Skellam and ΔNB models beat the ordered models. Here the Skellam model outperforms slightly the ΔNB model during most of the trading day.



(a) On 10th October 2008.



(b) On 30th April 2010.

Figure 6: Out-of-sample analysis: sequential predictive Bayes factor approximations, relative to the ordered normal model, for IBM (left) and KO (right) for the two periods.

5 Conclusions

We have reviewed and introduced dynamic models for high-frequency integer price changes. In particular, we have introduced the dynamic negative binomial difference model, referred to as the ΔNB model. We have developed a Markov chain Monte Carlo procedure (based on Gibbs sampling) for the Bayesian estimation of parameters in the dynamic Skellam and ΔNB models. Furthermore, we have demonstrated our estimation procedures for simulated data and for real data consisting of tick by tick transaction bid prices from NYSE stocks. We have compared the in-sample and out-of-sample performances of two classes of models, the ordered probit models and models based on integer distributions.

Our modeling framework opens several directions for future research. For instance, the ΔNB model has been defined by a time varying specification for the λ parameter in the ΔNB distribution, while the second parameter ν is kept constant over time. It can be of interest to investigate the impact of reversing this specification by considering a dynamic model for ν.

It would be also of interest to allow for a ΔNB distribution with a non-zero mean. This would allow to base our analysis on the non-centered parametrization of our state space model and hence to adopt the ancillarity-sufficiency interweaving strategy (ASIS) of Kastner and Frühwirth-Schnatter (2014) for the improvement of mixing the proposed sampler. This direction of improvement of the efficiency of our proposed sampler can also be considered for future research.

References

Aït-Sahalia, Y., J. Jacod, and J. Li (2012). Testing for jumps in noisy high frequency data. *Journal of Econometrics 168*, 207–222.

Aït-Sahalia, Y., P. A. Mykland, and L. Zhang (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics 160*, 160–175.

Alzaid, A. and M. A. Omair (2010). On the Poisson difference distribution inference and applications. *Bulletin of the Malaysian Mathematical Science Society 33*, 17–45.

Andersen, T. G. (2000). Some reflections on analysis of high-frequency data. *Journal of Business Economic Statistics 18(2)*, 146–153.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Realized Kernels in Practice: Trades and Quotes. *Econometrics Journal 4*, 1–32.

Barndorff-Nielsen, O. E., D. G. Pollard, and N. Shephard (2012). Integer-valued Lévy Processes and Low Latency Financial Econometrics. *Quantitative Finance 12*(4), 587–605.

Bos, C. (2008). Model-based estimation of high frequency jump diffusions with microstructure noise and stochastic volatility. TI Discussion Paper.

Boudt, K., J. Cornelissen, and S. Payseur (2012). Highfrequency: Toolkit for the analysis of Highfrequency financial data in R.

Brownlees, C. and G. Gallo (2006). Financial econometrics analysis at ultra-high frequency: Data handling concerns. *Compuational Statistics and Data Analysis 51*, 2232–2245.

Chakravarty, S., R. A. Wood, and R. A. V. Ness (2004). Decimals and liquidity: A study of the NYSE. *Journal of Financial Research 27*, 75–94.

Chib, S., F. Nardari, and N. Shephard (2002). Markov Chain Monte Carlo for stochastic volatility models. *Journal of Econometrics 108*, 281–316.

Chordia, T. and A. Subrahmanyam (1995). Market making, the tick size and payment-for-order-flow: Theory and evidence. *Journal of Business 68*, 543–576.

Cordella, T. and T. Foucault (1999). Minimum price variations, time priority and quote dynamics. *Journal of Financial Intermediation 8*, 141–173.

Czado, C. and S. Haug (2010). An ACD-ECOGARCH(1,1) model. *Journal of Financial Econometrics 8*, 335–344.

Dahlhaus, R. and J. C. Neddermeyer (2014). Online Spot Volatility-Estimation and Decompostion with Nonlinear Market Microstructure Noise Models. *Journal of Financial Econometrics 12*, 174–212.

Dayri, K. and M. Rosenbaum (2013). Large tick assets: implicit spread and optimal tick size. Working paper.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*, 603–616.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods* (Second Edition ed.). Oxford University Press.

Eisler, Z., J. P. Bouchaud, and J. Kockelkoren (2012). The price impact of order book events: market orders, limit orders and cancellations. *Quantitative Finance 12*, 1395–1419.

Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica 68*, 1–22.

Frühwirth-Schnatter, S., R. Frühwirth, L. Held, and H. Rue (2009). Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Statistics and Computing 19*, 479–492.

Frühwirth-Schnatter, S. and H. Wagner (2006). Auxiliary mixture sampling for parameter-driven models of time series of small counts with applications to state space modeling. *Biometrika 93*, 827–841.

Griffin, J. and R. Oomen (2008). Sampling retruns for realized varaiance calculations: tick time or transaction time? *Econometric Reviews 27*, 230–253.

Johnson, N. L., A. W. Kemp, and S. Kotz (2005). *Univariate Discrete Distributions* (Third Edition ed.). John Wiley and Sons.

Kastner, G. and S. Frühwirth-Schnatter (2014). Ancillarity-Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Estimation of Stochastic Volatility Models. *Computational Statistics & Data Analysis 76*, 408–423.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

Koopman, S. J., R. Lit, and A. Lucas (2017). Intraday Stochastic Volatility in Discrete Price Changes: the Dynamic Skellam Model. *Journal of the American Statistical Association 112*, 1490–1503.

Müller, G. and C. Czado (2009). Stochastic volatility models for ordinal-valued time series with application to finance. *Statistical Modelling 9*(1), 69–95.

O'Hara, M., G. Saar, and Z. Zhong (2014). Relative Tick Size and the Trading Environment. Working paper.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007). Stochastic volatility with leverage: fast likelihood inference. *Journal of Econometrics 140*, 425–449.

Poirier, D. J. (1973). Piecewise Regression Using Cubic Splines. *Journal of the American Statistical Association 68*, 515–524.

Roberts, G. O. and J. S. Rosenthal (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics 18*, 349–367.

Ronen, T. and D. G. Weaver (2001). "Teenies" anyone? *Journal of Financial Markets 4*, 231–260.

Rydberg, T. H. and N. Shephard (2003). Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics 1*, 2–25.

SEC (2012). Report to Congress on Decimalization. US Securities and Exchange Commission report.

Skellam, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society 109(3)*, 296.

Stefanos, D. (2015). Bayesian inference for ordinal-response state space mixed models with stochastic volatility. Working paper.

Stroud, J. R. and M. S. Johannes (2014). Bayesian modeling and forecasting of 24-hour high-frequency volatility: A case study of the financial crisis.

Weinberg, J., L. D. Brown, and R. S. J (2007). Bayesian forecasting of an imhomogenous Poisson process with application to call center data. *Journal of the American Statistical Association 102*, 1185–1199.

Ye, M. and C. Yao (2014). Tick Size Constrains, Market Structure, and Liquidity. Working paper.

APPENDICES

A Negative Binomial distribution

The probability mass function (pmf) of the NB distribution is given by

$$f(k;\nu,p)=\frac{\Gamma(\nu+k)}{\Gamma(\nu)\Gamma(k+1)}p^k(1-p)^\nu.$$

Its different parametrization can be obtained by denoting its mean by $\lambda=\nu\frac{p}{1-p}$, which implies $p=\frac{\lambda}{\lambda+\nu}$. We refer to this parametrization as $NB(\lambda,\nu)$. Then the pmf takes the following form

$$f(k;\lambda,\nu)=\frac{\Gamma(\nu+k)}{\Gamma(\nu)\Gamma(k+1)}\left(\frac{\lambda}{\nu+\lambda}\right)^k\left(\frac{\nu}{\nu+\lambda}\right)^\nu$$

and the variance is equal to $\lambda\left(1+\frac{\lambda}{\nu}\right)$. Then the dispersion index, or the variance-to-mean ratio, is equal to $\left(1+\frac{\lambda}{\nu}\right)>1$, which shows that the NB distribution is overdispersed. This means that there are more intervals with low counts and more intervals with high counts, compared to the Poisson distribution. The latter is nested in the NB distribution as the limiting case when $\nu\to\infty$.

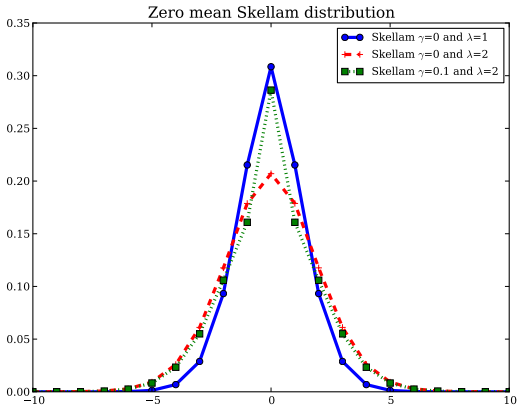
Alternatively, the NB distribution can be written as a Poisson-Gamma mixture. Let Y follow a Poisson distribution with the mean λU, where the heterogeneity parameter U has the unit mean and is Gamma-distributed, $U\sim\mathrm{Ga}(\nu,\nu)$, with the density of $\mathrm{Ga}(\alpha,\beta)$ given by $f(x;\alpha,\beta)=\frac{\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)}$. Then

$$\begin{aligned}f(k;\lambda,\nu)&=\int_0^\infty f_{\mathrm{Poisson}}(k;\lambda u)f_{\mathrm{Gamma}}(u;\nu,\nu)\mathrm{d}u\\&=\int_0^\infty\frac{(\lambda u)^k e^{-\lambda u}}{k!}\frac{\nu^\nu u^\nu e^{-\nu u}}{\Gamma(\nu)}\mathrm{d}u\\&=\frac{\lambda^k\nu^\nu}{k!\Gamma(\nu)}\int_0^\infty e^{-(\lambda+\nu)u}u^{k+\nu-1}\mathrm{d}u.\end{aligned}$$

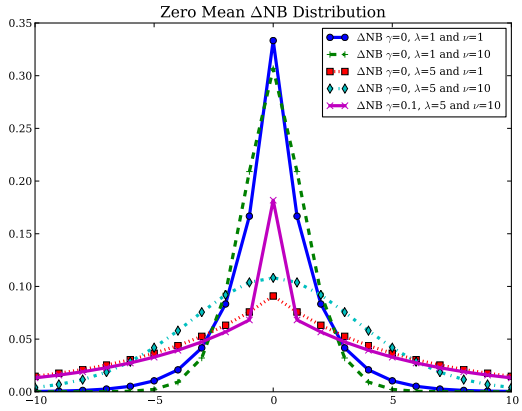
Substituting $(\lambda+\nu)u = s$, we get

$$
\begin{aligned}
&= \frac{\lambda^k \nu^\nu}{k!\Gamma(\nu)} \int_0^\infty e^{-s} \frac{s^{k+\nu-1}}{(\lambda+\nu)^{k+\nu-1}} \frac{1}{(\lambda+\nu)} \mathrm{d}s \\
&= \frac{\lambda^k \nu^\nu}{k!\Gamma(\nu)} \frac{1}{(\lambda+\nu)^{k+\nu}} \int_0^\infty e^{-s} s^{k+\nu-1} \mathrm{d}s \\
&= \frac{\lambda^k \nu^\nu}{k!\Gamma(\nu)} \frac{\Gamma(k+\nu)}{(\lambda+\nu)^{k+\nu}} \\
&= \frac{\Gamma(\nu+k)}{\Gamma(\nu)\Gamma(k+1)} \left(\frac{\lambda}{\nu+\lambda}\right)^k \left(\frac{\nu}{\nu+\lambda}\right)^\nu,
\end{aligned}
$$

which shows that $Y \sim NB(\lambda, \nu)$.



(a) The Skellam distribution with different parameters.



(b) The ΔNB distribution with different parameters.

B Daily volatility patterns

We want to approximate the function $f : \mathbb{R} \to \mathbb{R}$ with a continuous function which is built up from piecewise polynomials of degree at most three. Let the set $\Delta = \{k_0, \ldots, k_K\}$ denote the set of of knots k_j $j = 0, \ldots, K$. Δ is some times called a mesh on $[k_0, k_K]$. Let $y = \{y_0, \ldots, y_K\}$ where $y_j = f(x_j)$. We denote a cubic spline on Δ interpolating to y as $S_\Delta(x)$. $S_\Delta(x)$ has to satisfy the following conditions.

1. $S_\Delta(x) \in C^2\,[k_0, k_K]$.
2. $S_\Delta(x)$ coincides with a polynomial of degree at most three on the intervals $[k_{j-1}, k_j]$ for $j = 0, \ldots, K$.
3. $S_\Delta(x) = y_j$ for $j = 0, \ldots, K$.

Using condition 2 we know that the $S''_\Delta(x)$ is a linear function on $[k_{j-1}, k_j]$, which means that we can write $S''_\Delta(x)$ as

$$S''_\Delta(x) = \left[\frac{k_j - x}{h_j}\right] M_{j-1} + \left[\frac{x - k_{j-1}}{h_j}\right] M_j \quad \text{for} \quad x \in [k_{j-1}, k_j]\,,$$

where $M_j = S''_\Delta(k_j)$ and $h_j = k_j - k_{j-1}$. Integrating $S''_\Delta(x)$ and solving the integrating for the two integrating constants (using $S_\Delta(x) = y_j$) Poirier (1973) shows that we get

$$S'_\Delta(x) = \left[\frac{h_j}{6} - \frac{(k_j - x)^2}{2h_j}\right] M_{j-1} + \left[\frac{(x - k_{j-1})^2}{2h_j} - \frac{h_j}{6}\right] M_j + \frac{y_j - y_{j-1}}{h_j}, \quad \text{for} \quad x \in [k_{j-1}, k_j]$$

and

$$\begin{aligned} S_\Delta(x) =& \frac{k_j - x}{6h_j}\left[(k_j - x)^2 - h_j^2\right] M_{j-1} + \frac{x - k_{j-1}}{6h_j}\left[(x - k_{j-1})^2 - h_j^2\right] M_j \\ &+ \left[\frac{k_j - x}{h_j}\right] y_{j-1} + \left[\frac{x - k_{j-1}}{h_j}\right] y_j, \quad \text{for} \quad x \in [k_{j-1}, k_j] \end{aligned} \tag{A1}$$

In the above expressions only M_j for $j = 0, \ldots, K$ are unknown. We can use the continuity restrictions which enforce continuity at the knots by requiring that the derivatives are equal at the knots k_j for $j = 1, \ldots, K-1$, so that

$$\begin{aligned} S'_\Delta(k_j^-) &= h_j M_{j-1}/6 + h_j M_j/3 + (y_j - y_{j-1})/h_j, \\ S'_\Delta(k_j^+) &= -h_{j+1} M_j/3 - h_{j+1} M_{j+1}/6 + (y_{j+1} - y_j)/h_{j+1}, \end{aligned}$$

which yields $K - 1$ conditions

$$(1 - \lambda_j) M_{j-1} + 2M_j + \lambda_j M_{j+1} = \frac{6y_{j-1}}{h_j(h_j + h_{j+1})} - \frac{6y_j}{h_j h_{j+1}} + \frac{6y_{j+1}}{h_{j+1}(h_j + h_{j+1})},$$

where

$$\lambda_j = \frac{h_{j+1}}{h_j + h_{j+1}}.$$

Using two end conditions we have $K + 1$ unknowns and $K + 1$ equations and we can solve the linear equation system for M_j. Using the $M_0 = \pi_0 M_1$ and $M_K = \pi_K M_{K-1}$ end

conditions we can write

$$\Lambda = \begin{bmatrix} 2 & -2\ \pi_0 & 0 & \dots & 0 & 0 & 0 \\ 1\text{-}\lambda_1 & 2 & \lambda_1 & \dots & 0 & 0 & 0 \\ 0 & 1\text{-}\lambda_2 & 2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 2 & \lambda_{K-2} & 0 \\ 0 & 0 & 0 & \dots & 1\text{-}\lambda_{K-1} & 2 & \lambda_{K-1} \\ 0 & 0 & 0 & \dots & 0 & -2\ \pi_K & 2 \end{bmatrix},$$

$$\Theta = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \frac{6}{h_1(h_1+h_2)} & -\frac{6}{h_1 h_2} & \frac{6}{h_2(h_1+h_2)} & \dots & 0 & 0 & 0 \\ 0 & \frac{6}{h_2(h_2+h_3)} & -\frac{6}{h_2 h_3} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\frac{6}{h_{K-2}h_{K-1}} & \frac{6}{h_{K-1}(h_{K-2}+h_{K-1})} & 0 \\ 0 & 0 & 0 & \dots & \frac{6}{h_{K-1}(h_{K-1}+h_K)} & -\frac{6}{h_{K-1}h_K} & \frac{6}{h_K(h_{K-1}+h_K)} \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{bmatrix}$$

$m = (M_0, M_1, \dots, M_{K-1}, M_K)'$ and $y = (y_0, y_1, \dots, y_{K-1}, y_K)'$. The linear equation system is given by

$$\Lambda m = \theta y \tag{A2}$$

and the solution is

$$m = \Lambda^{-1}\Theta y \tag{A3}$$

Using this result and equation (A1) we can calculate

$$S_\Delta(\xi) = [S_\Delta(\xi_1), S_\Delta(\xi_2), \dots, S_\Delta(\xi_{N-1}), S_\Delta(\xi_N)]'.$$

Let us denote by P the $N \times (K+1)$ matrix in which the ith row $i = 1, \dots, N1$, given that $k_{j-1} \le \xi \le k_j$, can be written as

$$\underbrace{p_i}_{1\times(K+1)} = \left[\underbrace{0, \dots, 0}_{\text{first } j-2}, \frac{k_j - \xi_i}{6h_j}\left[(k_j - \xi_i)^2 - h_j^2\right], \frac{\xi_i - k_{j-1}}{6h_j}\left[(\xi_i - k_{j-1})^2 - h_j^2\right], \underbrace{0, \dots, 0}_{\text{last } K+1-j}\right].$$

Moreover, denote by Q the $N \times (K+1)$ matrix in which the ith row $i = 1, \dots, N$, given

that $k_{j-1} \leq \xi \leq k_j$, can be written as

$$\underbrace{q_i}_{1\times(K+1)} = \left[\underbrace{0,\ldots,0}_{\text{first } j-2}, \frac{k_j-\xi_i}{h_j}, \frac{\xi_i-k_{j-1}}{h_j}, \underbrace{0,\ldots,0}_{\text{last } K+1-j}\right].$$

Now using (A1) and (A3), we get

$$S_\Delta(\xi) = Pm + Qy = P\Lambda^{-1}\Theta y + Qy = (P\Lambda^{-1}\Theta + Q)y = \underbrace{W}_{N\times(K+1)}\underbrace{y}_{(K+1)\times 1},$$

where

$$W = P\Lambda^{-1}\Theta + Q.$$

In practical situations we might only know the knots but we do not know the spline values, which we observe with errors. In this case we have

$$s = S_\Delta(\xi) + \varepsilon = Wy + \varepsilon,$$

where $s = (s_1, s_2, \ldots, s_{N-1}, s_N)'$ and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{N-1}, \varepsilon_N)'$, with

$$\mathrm{E}(\varepsilon) = 0 \quad \text{and} \quad E(\varepsilon\varepsilon') = \sigma^2 I.$$

Notice that after fixing the knots we only have to estimate the value of the spline at the knots and this fully determines the shape of the spline. We can do this by a simple OLS regression

$$\hat{y} = (W^\top W)^{-1} W^\top s.$$

For the identification reasons we want

$$\sum_{j:\text{unique } \xi_j} S_\Delta(\xi_j) = \sum_{j:\text{unique } \xi_j} w_j y = w^* y = 0,$$

where w_i is the ith row of W and

$$w^* = \sum_{j:\text{unique } \xi_j} w_j.$$

To this end, a restriction can be enforced on one of the elements of y. This ensures that $E(s_t) = 0$, so that s_t and μ_h can be identified. If we drop y_K we can substitute

$$y_K = -\sum_{i=0}^{K-1} (w_i^*/w_K^*) y_i,$$

where w_i^* is the ith element of w^*. Substituting this into

$$\begin{aligned}\sum_{j:\text{unique } \xi_j} S_\Delta(\xi_j) &= \sum_{j:\text{unique } \xi_j} w_j y = \sum_{j:\text{unique } \xi_j} \sum_{i=0}^{K} w_{ji} y_i = \sum_{j:\text{unique } \xi_j} \sum_{i=0}^{K-1} w_{ji} y_i - w_{jK} \sum_{i=0}^{K-1} (w_i^*/w_K^*) y_i \\ &= \sum_{j:\text{unique } \xi_j} \sum_{i=0}^{K-1} (w_{ji} - w_{jK} w_i^*/w_K^*) y_i = \sum_{i=0}^{K-1} \sum_{j:\text{unique } \xi_j} (w_{ji} - w_{jK} w_i^*/w_K^*) y_i \\ &= \sum_{i=0}^{K-1} (w_i^* - w_K^* w_i^*/w_K^*) y_i = \sum_{i=0}^{K-1} (w_i^* - w_i^*) y_i = 0.\end{aligned}$$

Lets partition W in the following way

$$\underbrace{W}_{N\times(K+1)} = [\underbrace{W_{-K}}_{N\times K} : \underbrace{W_K}_{N\times 1}]$$

where W_{-K} is equal to the first K columns of W and W_K is the Kth column of W. Moreover

$$\underbrace{w^*}_{1\times(K+1)} = [\underbrace{w_{-K}^*}_{1\times K} : \underbrace{w_K^*}_{1\times 1}]$$

Finally, we can define

$$\underbrace{\widetilde{W}}_{N\times K} = \underbrace{W_{-K}}_{N\times K} - \frac{1}{w_K^*} \underbrace{W_K}_{N\times 1} \underbrace{w_{-K}^*}_{1\times K},$$

so that we obtain

$$s = S_\Delta(\xi) + \varepsilon = \underbrace{\widetilde{W}}_{N\times K} \underbrace{\tilde{y}}_{K\times 1} + \varepsilon.$$

C MCMC estimation of the ordered t-SV model

In this section, the t element vectors $(v_1, \ldots, v_t)$ containing time dependent variables for all time time periods, are denoted by v, the variable without a subscript.

C.1 Generating the parameters $x, \mu_h, \varphi, \sigma_\eta^2$ (Step 2)

Notice that conditional on $C = \{c_t, t = 1, \ldots, T\}$, r_t^* we have

$$2 \log r_t^* = \mu + s_t + x_t + \log \lambda_t + m_{c_t} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, v_{c_t}^2)$$

which implies the following state space form

$$\tilde{y}_t = \underbrace{\begin{bmatrix} 1 & w_t & 1 \end{bmatrix}}_{1\times(K+2)} \underbrace{\begin{bmatrix} \mu \\ \beta \\ x_t \end{bmatrix}}_{(K+2)\times 1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, v_{c_t}^2), \tag{A4}$$

$$\alpha_{t+1} = \underbrace{\begin{bmatrix} \mu \\ \beta \\ x_{t+1} \end{bmatrix}}_{(K+2)\times 1} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & I_K & 0 \\ 0 & 0 & \varphi \end{bmatrix}}_{(K+2)\times(K+2)} \underbrace{\begin{bmatrix} \mu \\ \beta \\ x_t \end{bmatrix}}_{(K+2)\times 1} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \eta_t \end{bmatrix}}_{(K+2)\times 1}, \eta_t \sim \mathcal{N}(0, \sigma_\eta^2), \tag{A5}$$

where

$$\underbrace{\begin{bmatrix} \mu \\ \beta \\ x_1 \end{bmatrix}}_{(K+2)\times 1} \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} \mu_0 \\ \beta_0 \\ 0 \end{bmatrix}}_{(K+2)\times 1}, \underbrace{\begin{bmatrix} \sigma_\mu^2 & 0 & 0 \\ 0 & \sigma_\beta^2 I_K & 0 \\ 0 & 0 & \sigma_\eta^2/(1-\varphi^2) \end{bmatrix}}_{(K+2)\times(K+2)} \right) \tag{A6}$$

and

$$\tilde{y}_t = 2\log r_t^* - \log \lambda_t - m_{r_{t1}}. \tag{A7}$$

First we draw φ, σ_η^2 from $p(\varphi, \sigma_\eta^2|\gamma, \nu, C, \tau, N, z_1, z_2, s, y)$. Notice that

$$p(\varphi, \sigma_\eta^2|\gamma, \nu, C, \tau, N, z_1, z_2, s, y) = p(\varphi, \sigma_\eta^2|\tilde{y}_t, C, N) \propto p(\tilde{y}_t|\varphi, \sigma_\eta^2, C, N)p(\varphi)p(\sigma_\eta^2),$$

where $\tilde{y}_t$ is defined above in equation (A11). The likelihood can be evaluated using standard Kalman filtering and prediction error decomposition (see e.g, Durbin and Koopman, 2012) taking advantage of fact that conditional on the auxiliary variables we have a linear Gaussian state space form given by equation (A8),(A9), (A10) and (A11). We draw from the posterior using an adaptive random walk Metropolis-Hastings step proposed by Roberts and Rosenthal (2009). Conditional on φ, σ_η^2 we draw μ_h, s and x from $p(\mu_h, s, x|\varphi, \sigma_\eta^2, \gamma, \nu, C, \tau, N, z_1, z_2, s, y)$, which is done simulating from the smoothed state density of the linear Gaussian state space model given by (A4),(A5), (A6) and (A7). We use the simulation smoother proposed by Durbin and Koopman (2002).

C.2 Generating γ (Step 3)

The conditional distribution for γ simplifies as follows

$$p(\gamma|\nu, \mu, \varphi, \sigma_\eta^2, x, s, C, y, r_t^*) = p(\gamma|\nu, h, y),$$

because given ν, h and y, γ does not depend on $C, \varphi, \sigma_\eta^2, r_t^*$. We further have that

$$p(\gamma|\nu, h, y) \propto p(y|\gamma, \nu, h)p(\gamma|\nu, h) = p(y|\gamma, \nu, h)p(\gamma),$$

as γ is independent from ν and h. Finally,

$$\begin{aligned}
p(y|\gamma, \nu, h)p(\gamma) =& \prod_{t=1}^{T} \Big\{ \gamma \mathbf{1}_{\{y_t=0\}} + (1-\gamma) \\
& \times \left[\mathcal{T}\left(\frac{y_t + 0.5}{\exp(h_t/2)}, \nu \right) - \mathcal{T}\left(\frac{y_t - 0.5}{\exp(h_t/2)}, \nu \right) \right] \Big\} \frac{\gamma^{a-1}(1-\gamma)^{b-1}}{\mathrm{B}(a,b)} \\
\propto & \prod_{t=1}^{T} \Big\{ \gamma^a (1-\gamma)^{b-1} \mathbf{1}_{\{y_t=0\}} + \gamma^{a-1}(1-\gamma)^b \\
& \times \left[\mathcal{T}\left(\frac{y_t + 0.5}{\exp(h_t/2)}, \nu \right) - \mathcal{T}\left(\frac{y_t - 0.5}{\exp(h_t/2)}, \nu \right) \right] \Big\},
\end{aligned}$$

where $\mathcal{T}(\cdot, \nu)$ is the Student's t density function with mean zero scale one and degree of freedom parameter ν. We sample from this posterior using an adaptive random walk Metropolis-Hastings sampler by Roberts and Rosenthal (2009).

C.3 Generating r^*

First, notice that the conditional distribution for r^* can be simplified as follows

$$p(r^*|\gamma, \nu, \mu, \varphi, \sigma_\eta^2, x, s, C, \lambda, y) = p(r^*|\gamma, h, \lambda, y) = \prod_{t=1}^{T} p(r_t^*|\gamma, h_t, \lambda_t, y_t).$$

Then, by the law of total probability we have

$$\begin{aligned}
p(r_t^*|\gamma, \nu, h_t, y_t) =& p(r_t^*|\gamma, \nu, h_t, \lambda_t, y_t, \text{zero}) p(\text{zero}|\gamma, h_t, \lambda_t, y_t) \\
& + p(r_t^*|\gamma, h_t, \lambda_t, y_t, \text{non-zero}) p(\text{non-zero}|\gamma, h_t, \lambda_t, y_t),
\end{aligned}$$

where $p(r_t^*|\gamma, h_t, \lambda_t, y_t, \text{zero})$ is a normal density with zero mean and variance $\lambda_t \exp(h_t)$ truncated to the interval $[y_t - 0.5, y_t + 0.5]$. If $y_t = 0$, then

$$\begin{aligned}
p(\text{zero}|\gamma, h_t, y_t = 0) &= \frac{p(\text{zero}, \gamma, h_t, y_t = 0)}{p(\gamma, h_t, y_t = 0)} = \frac{p(y_t = 0|\text{zero}, \gamma, h_t) p(\text{zero}|\gamma, h_t)}{p(y_t = 0|\gamma, h_t)} \\
&= \frac{1 \times \gamma}{\gamma + (1-\gamma)\left[\Phi\left(\frac{0.5}{\sqrt{\lambda_t}\exp(h_t/2)} \right) - \Phi\left(\frac{-0.5}{\sqrt{\lambda_t}\exp(h_t/2)}, \right) \right]}.
\end{aligned}$$

If $y_t = k \neq 0$, then

$$\begin{aligned} p(\text{zero}|\gamma, h_t, y_t = k) &= \frac{p(\text{zero}, \gamma, h_t, y_t = k)}{p(\gamma, h_t, y_t = k)} \\ &= \frac{p(y_t = k|\text{zero}, \gamma, h_t)p(\text{zero}|\gamma, h_t)}{p(y_t = k|\gamma, h_t)} = 0. \end{aligned}$$

Moreover $p(\text{non-zero}|\gamma, h_t, y_t) = 1 - p(\text{zero}|\gamma, h_t, y_t)$.

C.4 Generating ν and λ

To sample ν and λ we use the method by Stroud and Johannes (2014). We can decompose the posterior density as

$$p(\nu, \lambda|\gamma, \varphi, \sigma_\eta^2, h, C, y, r^*) = p(\nu, \lambda|h, r^*) = p(\lambda|\nu, h, r^*)p(\nu|h, r^*).$$

Note that we have to following mixture representation

$$r_t^* = \exp(h_t/2)\sqrt{\lambda_t}\varepsilon_t \quad \varepsilon_t \sim \mathcal{N}(0, 1) \quad \lambda_t \sim \text{IG}(\nu/2, \nu/2),$$

which implies

$$p(\nu|h, r^*) \propto \prod_{t=1}^{T} p\left(\frac{r_t^*}{\exp(h_t/2)}\middle|h_t, \nu\right) p(\nu),$$

where

$$p\left(\frac{r_t^*}{\exp(h_t/2)}\middle|h_t, \nu\right) \sim t_\nu(0, 1).$$

Combined with the prior $\nu \sim \mathcal{DU}(2, 128)$, this leads to the following posterior

$$p(\nu|h, r^*) \propto \prod_{t=1}^{T} p\left(\frac{r_t^*}{\exp(h_t/2)}\middle|h_t, \nu\right) = \prod_{t=1}^{T} g_{\nu^*}\left(\frac{r_t^*}{\exp(h_t/2)}\right) = \prod_{t=1}^{T} g_{\nu^*}(w_t),$$

where $w_t = r_t^*/\exp(h_t/2)$.

To avoid the computationally intense evaluation of these probabilities we can use a Metropolis-Hastings update. We can draw the proposal ν^* from the neighborhood of the current value $\nu^{(i)}$ using a discrete uniform distribution $\nu^* \sim DU(\nu^{(i)} - \delta, \nu^{(i)} + \delta)$ and accept with probability

$$\min\left\{1, \frac{\prod_{t=1}^{T} g_{\nu^*}(y_t)}{\prod_{t=1}^{T} g_{\nu^{(i)}}(y_t)}\right\},$$

where δ is chosen such that the acceptance rate is reasonable.

Finally, we have

$$p(\lambda|\nu,h,r^*)=\prod_{t=1}^{T}p(\lambda_t|\nu,h_t,r_t^*)\propto\prod_{t=1}^{T}p(r_t^*|\lambda_t,\nu,h_t)p(\lambda_t|\nu),$$

where

$$\begin{aligned}p\left(\frac{r_t^*}{\exp(h_t/2)}\middle|\lambda_t,\nu,h_t\right)&\sim\mathcal{N}(0,\lambda_t),\\ p(\lambda_t|\nu)&\sim\mathrm{IG}(\nu/2,\nu/2),\\ p(\lambda_t|\nu,h_t,r_t^*)&\sim\mathrm{IG}\left(\frac{\nu+1}{2},\frac{\nu+\left(\frac{r_t^*}{\exp(h_t/2)}\right)^2}{2}\right).\end{aligned}$$

D MCMC estimation of the dynamic ΔNB model

In this section, the t element vectors $(v_1,\ldots,v_t)$ containing time dependent variables for all time time periods, are denoted by v, the variable without a subscript. We discuss the algorithmic details for the ΔNB model and we note that these also apply to the dynamic Skellam model, except Step 4a (generating ν).

D.1 Generating x, s, μ_h, φ, σ_η^2 (Step 2)

Notice that conditional on $C=\{c_{tj},t=1,\ldots,T,j=1,\ldots,\min(N_t+1,2)\}$, τ, N,γ and s we have

$$-\log\tau_{t1}=\log(z_{t1}+z_{t2})+\mu_h+s_t+x_t+m_{c_{t1}}(1)+\varepsilon_{t1},\quad \varepsilon_{t1}\sim\mathcal{N}(0,v_{c_{t1}}^2(1))$$

and

$$-\log\tau_{t2}=\log(z_{t1}+z_{t2})+\mu_h+s_t+x_t+m_{c_{t2}}(N_t)+\varepsilon_{t2},\quad \varepsilon_{t2}\sim\mathcal{N}(0,v_{c_{t2}}^2(N_t)),$$

which implies the following state space form

$$\underbrace{\tilde{y}_t}_{\min(N_t+1,2)\times 1} = \underbrace{\begin{bmatrix} 1 & w_t & 1 \\ 1 & w_t & 1 \end{bmatrix}}_{\min(N_t+1,2)\times(K+2)} \underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_t \end{bmatrix}}_{(K+2)\times 1} + \underbrace{\varepsilon_t}_{\min(N_t+1,2)\times 1}, \quad \varepsilon_t \sim \mathcal{N}(0, \mathrm{H}_t), \tag{A8}$$

$$\alpha_{t+1} = \underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_{t+1} \end{bmatrix}}_{(K+2)\times 1} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & I_K & 0 \\ 0 & 0 & \varphi \end{bmatrix}}_{(K+2)\times(K+2)} \underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_t \end{bmatrix}}_{(K+2)\times 1} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \eta_t \end{bmatrix}}_{(K+2)\times 1}, \tag{A9}$$

where $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$ and

$$\underbrace{\begin{bmatrix} \mu_h \\ \beta \\ x_1 \end{bmatrix}}_{(K+2)\times 1} \sim \mathcal{N}\left(\underbrace{\begin{bmatrix} \mu_0 \\ \beta_0 \\ 0 \end{bmatrix}}_{(K+2)\times 1}, \underbrace{\begin{bmatrix} \sigma_\mu^2 & 0 & 0 \\ 0 & \sigma_\beta^2 I_K & 0 \\ 0 & 0 & \sigma_{eta}^2/(1-\varphi^2) \end{bmatrix}}_{(K+2)\times(K+2)} \right), \tag{A10}$$

with $\mathrm{H}_t = \mathrm{diag}(v_{c_{t1}}^2(1), v_{c_{t,2}}^2(N_t))$ and

$$\underbrace{\tilde{y}_t}_{\min(N_t+1,2)\times 1} = \begin{pmatrix} -\log \tau_{t1} - m_{r_{t1}}(1) - \log(z_{t1} + z_{t2}) \\ -\log \tau_{t2} - m_{r_{t2}}(N_t) - \log(z_{t1} + z_{t2}) \end{pmatrix} \tag{A11}$$

First we draw φ, σ_η^2 from $p(\varphi, \sigma_\eta^2 | \gamma, \nu, C, \tau, N, z_1, z_2, s, y)$. Notice that

$$p(\varphi, \sigma_\eta^2 | \gamma, \nu, C, \tau, N, z_1, z_2, s, y) = p(\varphi, \sigma_\eta^2 | \tilde{y}_t, C, N) \propto p(\tilde{y}_t | \varphi, \sigma_\eta^2, C, N) p(\varphi) p(\sigma_\eta^2), \tag{A12}$$

where $\tilde{y}_t$ is defined above in equation (A11). The likelihood can be evaluated using the standard Kalman filter and prediction error decomposition (see e.g, Durbin and Koopman, 2012) taking advantage of fact that conditional on the auxiliary variables we have a linear Gaussian state space form given by equation (A8)–(A11). We draw from the posterior using an adaptive random walk Metropolis-Hastings step proposed by Roberts and Rosenthal (2009). Conditional on φ, σ_η^2 we draw μ_h, s and x from $p(\mu_h, s, x | \varphi, \sigma_\eta^2, \gamma, \nu, C, \tau, N, z_1, z_2, s, y)$, which is done simulating from the smoothed state density of the linear Gaussian state space model given by (A8),(A9), (A10) and (A11). We use the simulation smoother proposed by Durbin and Koopman (2002).

D.2 Generating γ (Step 3)

Notice that we can simplify

$$p(\gamma|\nu,\mu_h,\varphi,\sigma_\eta^2,x,C,s,\tau,N,z_1,z_2,y) = p(\gamma|\nu,\mu_h,s,x,y) \tag{A13}$$

because given ν, λ and y, the variables C,τ,N,z_1,z_2 are redundant. We can then decompose

$$p(\gamma|\nu,\mu_h,s,x,y) \propto p(y|\gamma,\nu,\mu_h,s,x)p(\gamma|\nu,\mu_h,s,x) = p(y|\gamma,\nu,\mu_h,s,x)p(\gamma) \tag{A14}$$

as γ is independent from ν and $\lambda_t = \exp(\mu_h + s_t + x_t)$. Plugging in the formula for the likelihood and for the prior for γ yields

$$\begin{aligned}
p(y|\gamma,\nu,\mu_h,x)p(\gamma) &= \prod_{t=1}^{T}\left[\gamma\mathbf{1}_{\{y_t=0\}} + (1-\gamma)\left(\frac{\nu}{\lambda_t+\nu}\right)^{2\nu}\left(\frac{\lambda_t}{\lambda_t+\nu}\right)^{|y_t|}\frac{\Gamma(\nu+|y_t|)}{\Gamma(\nu)\Gamma(|y_t|)}\right. \\
&\times \left. F\left(\nu+y_t,\nu,y_t+1;\left(\frac{\lambda_t}{\lambda_t+\nu}\right)^2\right)\right]\frac{\gamma^{a-1}(1-\gamma)^{b-1}}{\mathrm{B}(a,b)} \\
&\propto \prod_{t=1}^{T}\left[\gamma^a(1-\gamma)^{b-1}\mathbf{1}_{\{y_t=0\}} + \gamma^{a-1}(1-\gamma)^b\left(\frac{\nu}{\lambda_t+\nu}\right)^{2\nu}\left(\frac{\lambda_t}{\lambda_t+\nu}\right)^{|y_t|}\right. \\
&\times \left.\frac{\Gamma(\nu+|y_t|)}{\Gamma(\nu)\Gamma(|y_t|)}\,F\left(\nu+y_t,\nu,y_t+1;\left(\frac{\lambda_t}{\lambda_t+\nu}\right)^2\right)\right].
\end{aligned}$$

We sample from this posterior using an adaptive random walk Metropolis-Hastings sampler.

D.3 Generating C,τ,N,z_1,z_2,ν (Step 4)

To start with, we decompose the joint posterior of C, τ, N, z_1, z_2 and ν into

$$\begin{aligned}
p(C,\tau,N,z_1,z_2,\nu|\gamma,\mu_h,\varphi,\sigma_\eta^2,s,x,y) =& p(C|\tau,N,z_1,z_2\gamma,p,\mu_h,\varphi,\sigma_\eta^2,s,x,y) \\
&\times p(\tau|N,z_1,z_2\gamma,\nu,\mu_h,\varphi,\sigma_\eta^2,s,x,y) \\
&\times p(N|z_1,z_2\gamma,\nu,\mu_h,\varphi,\sigma_\eta^2,s,x,y) \\
&\times p(z_1,z_2|\gamma,\nu,\mu_h,\varphi,\sigma_\eta^2,s,x,y) \\
&\times p(\nu|\gamma,\mu_h,\varphi,\sigma_\eta^2,s,x,y).
\end{aligned}$$

Generating ν (Step 4a)

Note that

$$\begin{aligned} p(\nu|\gamma, \mu_h, \varphi, \sigma_\eta^2, s, x, y) &= p(\nu|\gamma, \lambda, y) \\ &\propto p(\nu, \gamma, \lambda, y) \\ &= p(y|\gamma, \lambda, \nu)p(\lambda|\gamma, \nu)p(\gamma|\nu)p(\nu) \\ &= p(y|\gamma, \lambda, \nu)p(\lambda)p(\gamma)p(\nu) \\ &\propto p(y|\gamma, \lambda, \nu)p(\nu), \end{aligned}$$

where $p(y|\gamma, \lambda, \nu)$ is a product of zero inflated ΔNB probability mass functions.

We draw ν using a discrete uniform prior $\nu \sim DU(2, 128)$ and a random walk proposal in the following fashion as suggested by Stroud and Johannes (2014) for degree of freedom parameter of a t density. We can write the posterior as a multinomial distribution $p(\nu|\mu_h, x, z_1, z_2) \sim M(\pi_2^*, \ldots, \pi_{128}^*)$ with probabilities

$$\pi_\nu^* \propto \prod_{t=1}^{T} \left[\gamma \mathbb{I}_{\{y_t=0\}} + (1-\gamma) f_{\Delta\text{NB}}(y_t; \lambda_t, \nu)\right] = \prod_{t=1}^{T} g_\nu(y_t).$$

To avoid the computationally intense evaluation of these probabilities we can use a Metropolis-Hastings update. We can draw the proposal ν^* from the neighborhood of the current value $\nu^{(i)}$ using a discrete uniform distribution $\nu^* \sim DU(\nu^{(i)} - \delta, \nu^{(i)} + \delta)$ and accept with the probability

$$\min\left\{1, \frac{\prod_{t=1}^{T} g_{\nu^*}(y_t)}{\prod_{t=1}^{T} g_{\nu^{(i)}}(y_t)}\right\},$$

where δ is chosen such that the acceptance rate is reasonable.

Generating z_1, z_2 (Step 4b)

Notice that given γ, μ_h, s, x and y the elements of the vectors z_1 and z_2 are independent over time, so that their posterior distribution factorizes as follows

$$p(z_1, z_2|\gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s, x, y) = \prod_{t=1}^{T} p(z_{t1}, z_{t2}|\gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s_t, x_t, y_t).$$

Then we have for a single component

$$\begin{aligned} p(z_{t1}, z_{t2}|\gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s_t, x_t, y_t) \propto & p(z_{t1}, z_{t2}, \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s_t, x_t, y_t) \\ = & p(y_t|z_{t1}, z_{t2}, \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s_t, x_t) \\ & \times p(z_{t1}, z_{t2}|\gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s_t, x_t), \end{aligned}$$

which we express as

$$p(z_{t1}, z_{t2}|\gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s_t, x_t, y_t) \propto g(z_{t1}, z_{t2}) \frac{\nu^\nu z_{t1}^\nu e^{-\nu z_{t1}}}{\Gamma(\nu)} \frac{\nu^\nu z_{t2}^\nu e^{-\nu z_{t2}}}{\Gamma(\nu)},$$

where

$$g(z_{t1}, z_{t2}) = \left[\gamma \mathbf{1}_{\{y_t=0\}} + (1-\gamma) \exp\left[-\lambda_t(z_{t1}+z_{t2})\right] \left(\frac{z_{t1}}{z_{t2}}\right)^{\frac{y_t}{2}} I_{|y_t|}(2\lambda_t\sqrt{z_{t1}z_{t2}}) \right],$$

with $\lambda_t = \exp(\mu_h + s_t + x_t)$. We can carry out an independent MH step by sampling z_{1t}^* and z_{2t}^* from $\text{Ga}(\lambda_t, \nu)$ with the acceptance probability equal to $\min\left\{\frac{g(z_{1t}^*, z_{2t}^*)}{g(z_{t1}, z_{t2})}, 1\right\}$.

Generating N (Step 4c)

As described in Section 3.3.

Generating τ (Step 4d)

Notice that

$$p(\tau|N, z_1, z_2, \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, x, y) = p(\tau|N, \mu_h, z_1, z_2, s, x).$$

Moreover

$$\begin{aligned} p(\tau|\mu_h, z_1, z_2, s, x) &= \prod_{t=1}^{T} p(\tau_{1t}, \tau_{2t}|N_t, \mu_h, z_{t1}, z_{t2}, s_t, x_t), \\ &= \prod_{t=1}^{T} p(\tau_{1t}|\tau_{2t}, N_t, \mu_h, z_{t1}, z_{t2}, s_t, x_t) p(\tau_{2t}|N_t, \mu_h, z_{t1}, z_{t2}, s_t, x_t), \end{aligned}$$

where we can sample from $p(\tau_{2t}|N_t, \mu_h, z_{t1}, z_{t2}, s_t, x_t)$ using the fact that conditionally on N_t the arrival time τ_{2t} of the N_tth jump is the maximum of N_t uniform random variables and it has a $Beta(N_t, 1)$ distribution. The arrival time of the (N_t+1)th jump after 1 is exponentially distributed with intensity $\lambda_t(z_{t1}+z_{t2})$, hence

$$\tau_{1t} = 1 + \xi_t - \tau_{2t}, \quad \xi_t \sim \text{Exp}(\lambda_t(z_{t1}+z_{t2})).$$

Generating C (Step 4e)

Notice that

$$p(C|\tau, N, z_1, z_2, \gamma, \nu, \mu_h, \varphi, \sigma_\eta^2, s, x, y) = p(C|\tau, N, z_1, z_2, \nu, s, x).$$

Moreover,

$$p(C|\tau, N, z_1, z_2, \nu, s, x) = \prod_{t=1}^{T} \prod_{j=1}^{\min(N_t+1,2)} p(r_{tj}|\tau_t, N_t, \mu_h, z_{t1}, z_{t2}, s_t, x_t).$$

We can than sample c_{t1} from the following discrete distribution

$$p(c_{t1}|\tau_t, N_t, \mu_h, z_{t1}, z_{t2}, s_t, x_t) \propto w_k(1)\varphi(-\log\tau_{1t} - \log[\lambda_t(z_{t1}+z_{t2})], m_k(1), v_k^2(1)),$$

where $k = 1, \ldots, C(1)$. If $N_t > 0$ then draw r_{t2} from the discrete distribution

$$p(c_{t2}|\tau_t, N_t, \mu_h, z_{t1}, z_{t2}, s_t, x_t) \propto w_k(N_t)\varphi(-\log\tau_{1t} - \log[\lambda_t(z_{t1}+z_{t2})], m_k(N_t), v_k^2(N_t)),$$

for $k = 1, \ldots, C(N_t)$.

E Data cleaning and trade durations

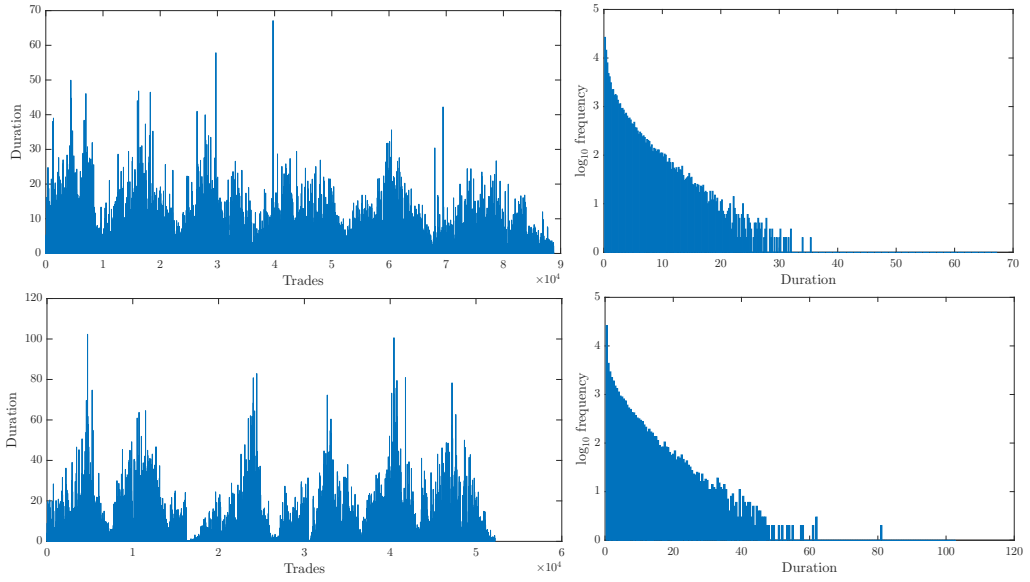
Tables 5–6 present the details of the data cleaning and aggregation procedure. Figure 8 presents the time series of the durations between the subsequent trades and their histograms (on the $\log_{10}$ scale for the frequencies), which are based on the cleaned data.

Table 5: Summary of the cleaning and aggregation procedure on the data from 3rd to 10th Oct 2008 for IBM and KO.

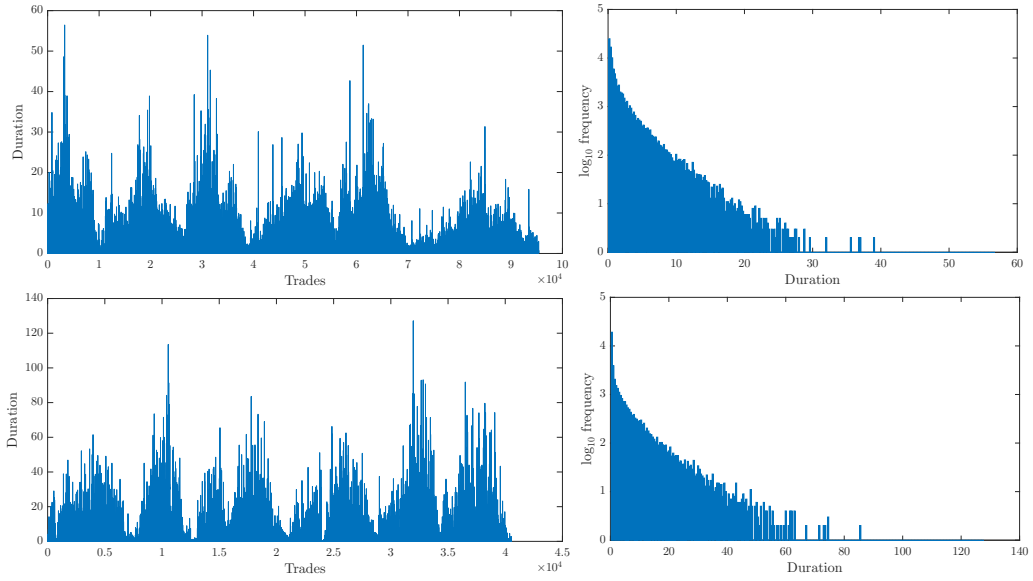
	IBM		KO	
	#	% dropped	#	% dropped
Raw quotes and trades	688 805		541 616	
Trades	128 589	81.33	126 509	76.64
Non missing price and volume	128 575	0.01	126 497	0.01
Trades between 9:30 and 16:00	128 561	0.01	126 484	0.01
Aggregated trades	89 517	30.37	96 482	23.72
Without outliers	88 808	0.79	95 398	1.12
Without opening trades	88 802	0.01	95 392	0.01

Table 6: Summary of the cleaning and aggregation procedure on the data from 23rd to 30th April 2010 for IBM and KO.

	IBM		KO	
	#	% dropped	#	% dropped
Raw quotes and trades	803 648		692 657	
Trades	53 346	93.36	41 184	94.05
Non missing price and volume	53 332	0.03	41 173	0.03
Trades between 9:30 and 16:00	53 324	0.02	41 164	0.02
Aggregated trades	52 406	1.72	40 573	1.44
Without outliers	52 199	0.39	40 548	0.06
Without opening trades	52 193	0.01	40 542	0.01



(a) IMB



(b) KO

Figure 8: Durations between trades and their $\log_{10}$ frequencies for IBM and KO, for the 2008 (top rows) and 2010 (bottom rows) samples, respectively.

F Intraday features, including overnight effects

In Figure 9 we present the estimated decompositions of the log volatility $h_t = \mu_h + s_t + x_t$ where we compare the component and signal estimates based on the model parameters that are estimated using all five days jointly and those from the model parameters that are estimated for each day separately. The motivation of this comparison is to verify whether the overnight effect, together with other intraday features, can be considered to be the same for each trading day or whether such features change from day to day. For our analysis of KO tick by tick transaction bid prices, we conclude that some features (such as intraday persistence) can be different from day to day but that the overall effects, including overnight effects, appear to be similar.

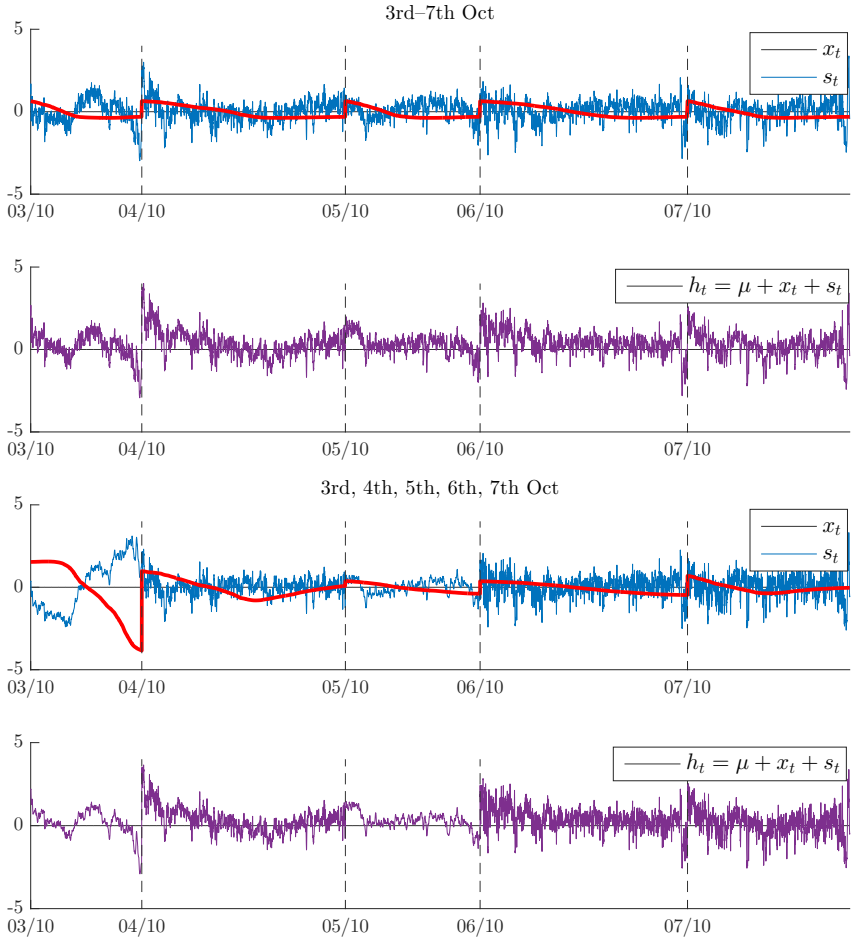


Figure 9: Volatility decomposition $h_t = \mu_h + s_t + x_t$ KO tick bid price returns for the 2008 data: ΔNB model parameters estimated based on the full sample (top two panels) and for each day separately (bottom two panels).